

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
12 September 2003 (12.09.2003)

PCT

(10) International Publication Number  
**WO 03/074679 A2**

(51) International Patent Classification<sup>7</sup>: **C12N**

(21) International Application Number: PCT/US03/06598

(22) International Filing Date: 3 March 2003 (03.03.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/360,843 1 March 2002 (01.03.2002) US  
60/384,197 29 May 2002 (29.05.2002) US

(71) Applicant (for all designated States except US): **XENCOR** [US/US]; 111 West Lemon Avenue, Monrovia, CA 91016 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **LAZAR, Gregory, Alan** [US/US]; 268 West Dryden Street, #114, Glendale, CA 91202 (US). **DESJARLAIS, John, Rudolph** [US/US]; 2096 E. Crary Street, Pasadena, CA 91104 (US). **MARSHALL, Shannon, Alicia** [US/US]; 1417 Bresee Avenue, Pasadena, CA 91104 (US). **DAHIYAT, Bassil** [US/US]; 3829 Luna Court, Altadena, CA 91001 (US).

(74) Agents: **SILVA, Robin, M.** et al.; Dorsey & Whitney LLP, 4 Embarcadero Center, Suite 3400, San Francisco, CA 94111 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,

LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Declaration under Rule 4.17:**

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: ANTIBODY OPTIMIZATION

(57) Abstract: The present invention relates to the use of computational screening methods to optimize the physico-chemical properties of antibodies, including stability, solubility, and antigen binding affinity.



WO 03/074679 A2

5

## ANTIBODY OPTIMIZATION

This application claims the benefit of the filing date of Serial No. 60/360,843, filed March 1, 2002 and Serial No. 60/384,197, filed May 29, 2002, both of which are expressly incorporated by reference in  
10 their entirety.

## FIELD OF THE INVENTION

The present invention relates to the use of computational screening methods to optimize the physico-  
15 chemical properties of antibodies, including stability, solubility, and antigen binding affinity.

## BACKGROUND OF THE INVENTION

Monoclonal antibodies are in widespread use as therapeutics, diagnostics, and research reagents.  
20 As therapeutics, antibodies are used to treat a variety of conditions including cancer, autoimmune diseases, and cardiovascular disease. There are currently over ten approved antibody products on the US market, with over a hundred in development. Despite such acceptance and promise, there remains significant need for optimization of the structural and functional properties of antibodies.

25 The physical and chemical properties of antibody therapeutics significantly determine their performance during development, manufacturing, and clinical use. Antibodies may suffer from the stability and solubility issues similar to all proteins. Since fully developed antibody therapeutics require high levels of stability and solubility in order to retain activity through purification, formulation, storage, and administration, there is a need for effective methods to optimize antibody properties.  
30 Antibodies may be exposed to a variety of stresses, for example changes in temperature or pH, that may cause protein unfolding, destroy activity, or make the protein sensitive to proteolytic degradation. Proteins may be reengineered such that structure and activity are substantially more robust with respect to such stresses, for example, by optimizing intramolecular and interdomain interactions and by altering protease recognition sites.

35 Solubility is also of critical importance to antibody efficacy. Antibodies are typically formulated and administered at high concentration, conditions under which antibodies may form aggregates. Aggregates typically have poor activity and bioavailability, and are associated with increased immunogenicity. Solubility may also dictate which routes of administration are feasible. In many  
40 cases, antibody therapeutics have been limited to intravenous administration, because the antibody is not sufficiently soluble to allow formulation of an effective dose in the small volumes that are used for alternate routes of administration. In most cases, solubility obstacles have been considered as formulation problems that may be surmounted with exhaustive protein chemistry effort. However, such methods are inefficient, inconsistent, and time-consuming, often failing to yield soluble protein

5 even following a significant expenditure of resources. Engineering approaches are beginning to emerge for the generation of soluble proteins; for example, in some cases solubility may be improved by replacing solvent exposed nonpolar residues with structurally compatible polar residues.

10 Another property of antibodies that frequently demands optimization is antigen-binding affinity. The binding affinity of an antibody for its biological target is a critical parameter for therapeutic efficacy. One particular case in which higher affinity is often sought is following humanization, herein defined as the reengineering of nonhuman antibodies to be more human-like in sequence. Humanization is carried out to reduce the immunogenicity of antibody therapeutics, but often results in loss of binding affinity for antigen. Regaining this affinity is typically desired during drug development. The main  
15 approach for enhancement of antigen affinity, herein referred to as affinity maturation, involves the engineering of mutations at positions that either directly contact antigen or indirectly influence binding. The demand for increased affinity for antigen is not, however, limited to humanization. Affinity maturation is frequently desired for therapeutic antibodies in general, whether they are derived from human, humanized, chimeric, or nonhuman sources.

20 Strategies for antibody optimization are sometimes carried out using random mutagenesis. In these cases positions are chosen randomly, or amino acid changes are made using simplistic rules. For example all residues may be mutated to alanine, referred to as alanine scanning. This can be used, for example, to map the antigen binding residues of an antibody (Kelley *et al.*, 1993, *Biochemistry*  
25 32:6828-6835; Vajdos *et al.*, 2002, *J. Mol. Biol.* 320:415-428). The high level of sequence and structural similarity and large amount of sequence and structural information enable sequence-based methods of optimization. For example, sequence analysis has allowed significant characterization of the determinants of antibody stability and solubility (Ewert *et al.*, 2003, *J. Mol. Biol.* 325:531-553; Ewert *et al.*, 2003, *Biochemistry* 42:1517-1528), and can enable sequence-based methods of affinity  
30 maturation (see, US 2003/0022240A1 and US 2002/0177170A1, both hereby incorporated by reference). Sequence and structural information can be coupled with site-directed mutagenesis to engineer antibodies with enhanced biophysical properties (Worn & Plückthun, 2001, *J. Mol. Biol.* 305:989-1010; Wirtz & Steipe, 1999, *Protein Sci.* 8:2245-2250). More sophisticated engineering approaches for implementing antibody optimization strategies employ selection methods to screen  
35 higher levels of sequence diversity. As is well known in the art, there are a variety of selection technologies which may be used for such approaches, including, for example, display technologies such as phage display, ribosome display, yeast display, and the like. Selection methods coupled with random or rational mutagenesis have found utility for optimizing antibody stability (Jung *et al.*, 1999, *J. Mol. Biol.* 294:163-180) and particularly for affinity maturation (Wu *et al.*, 1999, *J. Mol. Biol.* 294:151-  
40 162; Schier *et al.*, 1996, *J. Mol. Biol.* 255:28-43).

Despite some success, these current engineering strategies for antibody optimization suffer from three main obstacles. First, the level of sequence diversity that is wanted or needed can dramatically exceed that which is accessible by these technologies. The number of possible protein sequences

grows exponentially with the number of positions that are randomized. Practical considerations including experimental and physical constraints such as transformation efficiency, instrumentation limits, and the like can significantly limit library size. Even for methods capable of screening large combinatorial libraries, this presents an obstacle. For example, the upper limit of diversity accessible by phage display is approximately  $10^9$ , which limits mutations to 7 positions if a fully random (all 20 amino acids) library is used.

A second limitation of current antibody engineering efforts is that experimental screens used to assess the fitness of antibody variants are not efficient, and therefore engineering optimized antibodies can be time- and resource- intensive, with no guarantee of success. Nor do current experimental screens always have the capacity to be implemented as a selection. For example, antibody stability is not a property that is readily selected for using a display technology. Screening for more stable antibodies would require purifying individual variants and determining their thermodynamic stability using time consuming biophysical methods.

A final limitation of current antibody engineering efforts is that constraints on proteins are not distinct. Instead, the determinants of antibody stability, solubility, and affinity for antigen are overlapping and the interactions that contribute to these properties are related. Thus, affinity maturation of an antibody may result in decreased stability, and optimization of an antibody's solubility may cause a loss in affinity for its antigen. This issue has important ramifications for antibody engineering because current experimental antibody optimization methods are poorly suited for simultaneous optimization of multiple, related properties. Consequently, a large portion of the candidates in experimental libraries are unsuitable. For example, a large fraction of sequence space encodes unfolded, misfolded, incompletely folded, partially folded, or aggregated proteins. Even among sequences that are folded and active, many will be less active, less soluble, or less stable than the wild type protein. In effect, current antibody engineering efforts generate experimental libraries that are composed of a large amount of "wasted" sequence space. More significantly, the probability of finding a suitable sequence decreases dramatically as the number of properties that are considered increases. Thus, there is a need for computational screening methods to optimize the physico-chemical properties of antibodies, including stability, solubility, and antigen binding affinity.

## SUMMARY OF THE INVENTION

The present invention provides methods of computational screening that may be applied to enhance the stability of antibodies, the solubility of antibodies, and the affinity of antibodies for antigen.

More specifically, the present invention discloses a method for optimizing at least one physico-chemical property of an antibody, wherein the method is executed by a computer under the control of a program, and the computer including a memory for storing said program, said method comprising the steps of: a. receiving a template antibody structure; b. selecting at least one variable position

5 which belongs to said template antibody structure; c. selecting at least one amino acid to be considered at said variable positions; d. analyzing the interaction of each of said amino acids at each variable position with at least part of the remainder of said antibody, including said amino acids at other variable positions; and e. identifying a set of at least one antibody sequence with at least one optimized physico-chemical property.

10 The method of the present invention also optionally includes generating a library from the set of at least one antibody sequence and experimentally screening the library.

15 Computational screening methods have demonstrated their utility and success for the optimization of a broad array of protein properties. Application of these methods to antibodies represents a significant improvement because there are well known and established engineering strategies that are uniquely suited to antibodies. Computational screening is a hypothesis-driven method for engineering proteins, and thus the validity of the employed design strategies are critical to success. The application of these established engineering strategies as computational screening design strategies is not necessarily straightforward. However, as will be provided in detail, a number of aspects and parameters of the computational screening method may be adjusted to enable implementation of established antibody engineering strategies. Because all antibodies share a common structural template and high sequence similarity, and because of the enormous amount of sequence and structural information available, successful design strategies for the use of computational screening to optimize antibody stability, solubility, and affinity for antigen are broadly applicable to the entire family of antibodies. Finally, antibodies are often comprised of multiple similar domains. As a result, computational screening methods are uniquely modular for antibodies, that is to say that optimizations can be applied in an additive manner to engineer antibodies with a breadth of simultaneously enhanced functional and biophysical properties in multiple structural regions.

30 Computational screening methods of the present invention overcome the limitations of current antibody engineering methods. These methods are capitalizing on enormous recent advances in understanding of protein structure and function, substantial increases in the availability of high-resolution structures, and dramatic improvements in computing power. These methods offer a mechanism to explore sequence combinations that extend far beyond natural diversity, up to  $10^{50}$  or more sequences. Computational screening also enables the exploration of combinatorial complexity in the absence of experimentally selectable function, and thus biophysical properties such as stability and solubility, which are difficult to screen or select for, may be rationally screened in silico. Finally, computational screening methods offer the ability to algorithmically couple multiple constraints for simultaneous optimization of several protein properties. Thus experimental libraries that are designed using computational screening are composed primarily of productive sequence space. Computational screening may enrich experimental libraries with quality diversity, whether such experimental libraries are small such that members may be screened individually, or they are large such that selection

5 methods are required for screening. As a result, computational screening increases the chances of identifying antibodies that are broadly optimized for stability, solubility, and affinity for antigen.

An additional benefit of computational screening methodology is that it is hypothesis driven (dash here). Thus successful strategies may be reapplied to antibodies as a whole, saving discovery cost and time. This is particularly relevant for antibodies because all antibodies share a common structural template and high sequence similarity, and because of the enormous amount of sequence and structural information available.

15 It is an object of the present invention to provide design strategies for the application of computational screening methods to enhance the stability of antibodies, to enhance the solubility of antibodies, and to affinity mature antibodies. Said design strategies describe the theoretical and/or experimental basis for their use, how the choice of variable positions and amino acids considered at those positions are carried out for their implementation, and ways in which experimental and sequence information may be used.

20 It is a further object of the present invention to provide computational methods for the application of computational screening methods to enhance the stability of antibodies, to enhance the solubility of antibodies, and to affinity mature antibodies. These computational methods describe a broad array of scoring functions, optimization algorithms, and the like for implementing computer programs to optimize antibodies. The computational methods further describe ways by which computational output may be used to generate experimental libraries of variants for experimental validation.

30 It is another object of the present invention to provide experimental methods for the application of computational screening technology to enhance the stability of antibodies, to enhance the solubility of antibodies, and to affinity mature antibodies. The experimental methods describe a broad array of molecular biology, protein production, and screening techniques that may be used to experimentally validate antibody variants that have been optimized for improved properties using computational screening methods.

35 In accordance with the objects outlined above, the present invention provides computational screening methods to optimize antibodies.

#### BRIEF DESCRIPTION OF THE DRAWINGS

40 Figure 1. Antibody structure and function. Shown is a model of a full-length human IgG1 antibody, constructed by combining the structure of the Campath Fab fragment (pdb accession code 1CE1), with the structure of the human IgG1 Fc region (pdb accession code 1DN2). The antibody is a homodimer of heterodimers, made up of two light chains and two heavy chains. The Ig domains that comprise the antibody are labeled, and include V<sub>L</sub> and C<sub>L</sub> for the light chain, and V<sub>H</sub>, Cγ1 (Cγ1),

5 Cgamma2 (C $\gamma$ 2), and Cgamma3 (C $\gamma$ 3) for the heavy chain. Antibody regions relevant to the discussion are also labeled, including the variable region (Fv), the Fab region, and the Fc region. The regions which bind molecules or proteins relevant to the present invention are indicated, including the antigen binding site in the variable region, and the Fc region which binds Fc $\gamma$ Rs, FcRn, C1q, and proteins A and G. Campath is a registered trademark in the US of Burroughs Wellcome.

10 Figures 2a and 2b. Human germ line sequences and aligned antibody sequences. The sequences which are known to encode the human heavy chain variable region (V<sub>H</sub>) and the human kappa light chain variable region (V<sub>L</sub>) are shown aligned with four relevant antibody sequences. The germ line sequences were obtained from the IMGT database (IMGT, the international ImMunoGeneTics information system®; [imgt.cines.fr](http://imgt.cines.fr)), and aligned and numbered according to the numbering scheme of Chothia (Chothia *et al.*, 1992, *J. Mol. Biol.* 227, 776-798, 799-817; Tomlinson *et al.*, 1995, *EMBO J.* 14:4628-4638; Williams *et al.*, 1996, *J. Mol. Biol.* 264:220-232; Al-Lazikani *et al.*, 1997, *J. Mol. Biol.* 273, 927-948; Chothia *et al.*, 1998, *J. Mol. Biol.* 278, 457-479; all of which are herein expressly incorporated by reference). The regions of the variable region are indicated above the numbering, and these include framework regions 1 through 3 (FR1, FR2, and FR3) and the complementarity determining regions (CDRs) 1 through 3 (CDR1, CDR2, and CDR3). As is known in the art, V<sub>H</sub> CDR3 is not a part of the V<sub>H</sub> germ line and V<sub>L</sub> CDR3 is encoded only up to Chothia position 95 in the V<sub>L</sub> kappa germ line. Positions that make up CDRs are underlined. The germ line chains are grouped into 7 subfamilies for both V<sub>H</sub> and V<sub>L</sub>, as is known in the art, and these subfamilies are grouped together and separated by a blank line. Four antibody sequences used in the examples of the present invention, listed by their pdb accession codes and underlined, are shown below the subfamily to which they are closest in sequence. These sequences were aligned using the alignment program BLAST. The most similar germ line sequences to these four antibodies, as determined by this alignment analysis, are shown in parentheses next to the antibody code. The most similar germ line V<sub>H</sub> chains to the four antibodies are VH\_3-74 for D3H44 (1JPT), VH\_3-66 for Herceptin (1FVC), VH\_4-59 and VH\_3-72 for Campath (1CE1), and VH\_7-4-1 for rhumAb VEGF (1CZ8). The most similar germ line V<sub>L</sub> chains to the four antibodies are VLk\_1D-3 for D3H44 (1JPT), VLk\_1D-3 for Herceptin (1FVC), VLk\_1D-33 for Campath (1CE1), and VLk\_1D-33 for rhumAb VEGF (1CZ8). Herceptin is a registered trademark in the US owned by Genentech, Inc.

35 Figure 3. Antibody structures relevant to the presented examples. The seven antibody structures used in the present invention are listed. For each antibody is listed the target antigen, the source, the pdb accession code, whether the structure is a complex of the antibody with antigen (bound) or is uncomplexed (unbound), the resolution, and the reference.

40 Figure 4. Campath V<sub>H</sub> domain stabilization. The large central figure shows the Campath V<sub>H</sub> domain from 1CE1 as a gray ribbon diagram, with Example 1 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure (from Figure 1) with the relevant domain highlighted by a box.

5

Figures 5a, 5b, and 5c. Campath V<sub>H</sub> domain stabilization. Figure 5a shows the results of the computational screening calculations described in Example 1. Column 1 lists the heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core classification are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT Campath amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figures 5b and 5c show experimental libraries derived from the computational screening results, as described in Example 1. Column 1 lists variable positions and column 2 shows amino acid substitutions that are included in the experimental library. Figure 5c is represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the library, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

Figure 6. Campath V<sub>L</sub> domain stabilization. The large central figure shows the Campath V<sub>L</sub> domain from 1CE1 as a gray ribbon diagram, with Example 2 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant domain highlighted by a box.

Figures 7a and 7b. Campath V<sub>L</sub> domain stabilization. Figure 7a shows the results of the computational screening calculations described in Example 2. Column 1 lists the light (L) chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core and Boundary classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT Campath amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position which are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figure 7b shows an experimental library derived from the computational screening results, as described in Example 2. Column 1 lists variable positions and column 2 shows amino acid substitutions which are included in the experimental library. The library is represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the library, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

- 5 Figure 8. Campath V<sub>H</sub> C<sub>γ</sub>1 domain stabilization. The large central figure shows the Campath V<sub>H</sub> C<sub>γ</sub>1 domain from 1CE1 as a gray ribbon diagram, with Example 3 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant domain highlighted by a box.
- 10 Figures 9a and 9b. Campath V<sub>H</sub> C<sub>γ</sub>1 domain stabilization. Figure 9a shows the results of the computational screening calculations described in Example 3. Column 1 lists the heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core and Boundary classifications are described in the section entitled
- 15 "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT Campath amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figure 9b shows an experimental library
- 20 derived from the computational screening results, as described in Example 3. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The library is represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the library, that is the total number of defined
- 25 sequences of which it is composed, is shown in the bottom row.

Figure 10. Fc V<sub>H</sub> C<sub>γ</sub>2 domain stabilization. The large central figure shows the Fc V<sub>H</sub> C<sub>γ</sub>2 domain from 1DN2 as a gray ribbon diagram, with Example 4 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant

30 domain highlighted by a box.

- Figures 11a and 11b. Fc V<sub>H</sub> C<sub>γ</sub>2 domain stabilization. Figure 11a shows the results of the computational screening calculations described in Example 4. Column 1 lists the heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of
- 35 amino acids belonging to the Core and Boundary classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT Campath amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output.
- 40 Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figure 11b shows an experimental library derived from the computational screening results, as described in Example 4. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The library is represented combinatorially, that is the explicit library is the combination of each

5 possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the library, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

10 Figure 12. Fc V<sub>H</sub> C<sub>γ</sub>3 domain stabilization. The large central figure shows the Fc V<sub>H</sub> C<sub>γ</sub>3 domain from 1DN2 as a gray ribbon diagram, with Example 5 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant domain highlighted by a box.

15 Figures 13a and 13b. Fc V<sub>H</sub> C<sub>γ</sub>3 domain stabilization. Figure 13a shows the results of the computational screening calculations described in Example 5. Column 1 lists the heavy chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core and Boundary classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT Fc amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figure 13b shows an experimental library derived from the computational screening results, as described in Example 5. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The library is represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the library, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

30

Figure 14. rhumAb VEGF V<sub>H</sub>/V<sub>L</sub> interface stabilization. The large central figure shows the rhumAb VEGF V<sub>H</sub> and V<sub>L</sub> domains from 1CZ8 as black and gray ribbons respectively, with Example 6 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

35

Figures 15a, 15b, and 15c. rhumAb VEGF V<sub>H</sub>/V<sub>L</sub> interface stabilization. Figures 15a and 15b show the results of the computational screening calculations described in Example 6. Column 1 lists the light (L) and heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core and Boundary classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT rhumAb VEGF amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the

40

5 number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figure 15c shows an experimental library derived from the computational screening results, as described in Example 6. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The library is represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable  
 10 position with all other possible amino acid substitutions at all other positions. The complexity of the library, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

Figures 16a and 16b. Sequence alignment of rhumAb VEGF variable region with the human variable  
 15 region germ line. The rhumAb VEGF  $V_H$  and  $V_L$  sequences are shown aligned with the sequences that encode the human  $V_H$  (Figure 16a) and  $V_L$  (Figure 16b) germ line. The germ line sequences were obtained from the IMGT database, and numbered according to the numbering scheme of Chothia. The regions of the variable region are indicated above the numbering, and these include framework regions 1 through 3 (FR1, FR2, and FR3) and the complementarity determining regions  
 20 (CDRs) 1 through 3 (CDR1, CDR2, and CDR3). Positions that make up CDRs are underlined. The 7 germ line subfamilies for  $V_H$  and  $V_L$  are grouped together and separated by a blank line. The rhumAb VEGF  $V_H$  and  $V_L$  sequences were aligned to the germ line sequences using the alignment program BLAST. rhumAb VEGF  $V_H$  is most similar to the germ line chain VH\_7-4-1, and rhumAb VEGF  $V_L$  is most similar to the germ line chain VLk\_1D-33. The rhumAb VEGF  $V_H$  and  $V_L$  sequences are  
 25 indicated by the underlined pdb accession code 1CZ8, and shown below the subfamily to which they are closest in sequence. Amino acids at variable positions for Example 6 design calculations are shown in bold in the 1CZ8 and the germ line sequences.

Figures 17a and 17b. rhumAb VEGF sequence-guided  $V_H/V_L$  interface stabilization. Figure 17a  
 30 shows the results of the computational screening calculations described in Example 6. Rows 1 through 5 list the chain (L, light chain or H, heavy chain), variable positions as defined in the 1CZ8 structure and the according to the Chothia numbering scheme, amino acids considered at those positions as obtained from Figures 16a and 16b, and the amino acid at each position in the WT rhumAb VEGF sequence. "All" or "All 20" means that all 20 amino acids are considered at the  
 35 variable position. The rows that follow list the amino acid identity at variable positions for the lowest energy sequence from each cluster group, as described in Example 6. Figure 17a is similar to Figure 17b except that all the listed sequences are the set of sequences make up cluster group 5.

Figure 18. Herceptin  $V_H/V_L$  interface stabilization. The large central figure shows the Herceptin  $V_H$   
 40 and  $V_L$  domains from 1FVC as black and gray ribbons respectively, with Example 7 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

5 Figures 19a, 19b, 19c, and 19d. Herceptin  $V_H/V_L$  interface stabilization. Figures 19a and 19c show the results of the computational screening calculations described in Example 7. Column 1 lists the light (L) and heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core, Surface, and Boundary classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT Herceptin amino acid identity at each variable position.

10 Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figures 19b and 19d show experimental libraries derived from the computational screening results, as described in Example 7. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The libraries are represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the libraries, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

Figure 20. rhumAb VEGF  $C_L/C\gamma 1$  interface stabilization. The large central figure shows the VEGF  $C_L$  and  $C\gamma 1$  domains from 1CZ8 as black and gray ribbons respectively, with Example 8 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

Figures 21a and 21b. rhumAb VEGF  $C_L/C\gamma 1$  interface stabilization. Figure 21a shows the results of the computational screening calculations described in Example 8. Column 1 lists the light (L) and heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT rhumAb VEGF amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figure 21b shows an experimental library derived from the computational screening results, as described in Example 8. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The libraries are represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the libraries, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

- 5 Figure 22. Fc C $\gamma$ 3/C $\gamma$ 3 interface stabilization. The large central figure shows the Fc C $\gamma$ 3 domains from 1DN2 as gray ribbons, with Example 9 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.
- 10 Figures 23a and 23b. Fc C $\gamma$ 3/C $\gamma$ 3 interface stabilization. Figure 23a shows the results of the computational screening calculations described in Example 9. Column 1 lists the heavy chain variable positions. Chains A and B are the two symmetrical C $\gamma$ 3 domains in the 1DN2 structure. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core classifications are described in the section entitled "Selection of Amino Acids to
- 15 be Considered at Each Position". Column 3 lists the WT Fc amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that
- 20 amino acid at that variable position. Figure 23b shows an experimental library derived from the computational screening results, as described in Example 9. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The libraries are represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other
- 25 positions. The complexity of the libraries, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

Figure 24. Campath solubility optimization. The large central figure shows the Campath Fab fragment from 1CE1 as a gray ribbon diagram, with Example 10 variable position residues

30 represented as black ball and sticks. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

Figures 25a and 25b. Campath solubility optimization. Figure 25a shows the results of the computational screening calculations described in Example 10. Column 1 lists the heavy (H) and light

35 (L) chain variable positions. Column 2 lists the wild type amino acid identity at each variable position. The remaining 20 columns indicate which of the 20 natural amino acids are favorable substitutions for each variable position according to the computational screening calculations. The presence of an amino acid in its column for a variable position indicates that the amino acid is within 1 unit of energy of the lowest energy substitution. Figure 25b shows an experimental library derived from the

40 computational screening results, as described in Example 10. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The library is represented combinatorially, i.e. the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other

5 positions. The complexity of the library, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

Figure 26. rhumAb VEGF solubility optimization. The large central figure shows the rhumAb VEGF Fab fragment from 1CZ8 as a gray ribbon diagram, with Example 11 variable position residues  
10 represented as black ball and sticks. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

Figures 27a and 27b. rhumAb VEGF solubility optimization. Figure 27a shows the results of the computational screening calculations described in Example 11. Column 1 lists the heavy (H) and light  
15 (L) chain variable positions. Column 2 lists the wild type amino acid identity at each variable position. The remaining 20 columns indicate which of the 20 natural amino acids are favorable substitutions for each variable position according to the computational screening calculations. The presence of an amino acid in its column for a variable position indicates that the amino acid is within 1 unit of energy of the lowest energy substitution. Figure 27b shows an experimental library derived from the  
20 computational screening results, as described in Example 11. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The library is represented combinatorially, i.e. the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the library, that is the total number of defined sequences of which it is  
25 composed, is shown in the bottom row.

Figure 28. Herceptin solubility optimization. The large central figure shows the Herceptin scFv fragment from 1FVC as a gray ribbon diagram, with Example 12 variable position residues  
30 represented as black ball and sticks. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

Figures 29a and 29b. Herceptin solubility optimization. Figure 29a shows the results of the computational screening calculations described in Example 12. Column 1 lists the heavy (H) and light  
(L) chain variable positions. Column 2 lists the wild type amino acid identity at each variable position.  
35 The remaining 20 columns indicate which of the 20 natural amino acids are favorable substitutions for each variable position according to the computational screening calculations. The presence of an amino acid in its column for a variable position indicates that the amino acid is within 1 unit of energy of the lowest energy substitution. Figure 29b shows an experimental library derived from the computational screening results, as described in Example 12. Column 1 lists variable positions, and  
40 column 2 shows amino acid substitutions that are included in the experimental library. The library is represented combinatorially, i.e. the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the library, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

5

Figure 30. Fc solubility optimization. The large central figure shows the Fc region from 1DN2 as a gray ribbon diagram, with Example 13 variable position residues represented as black ball and sticks. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

10

Figures 31a and 31b. Fc solubility optimization. Figure 31a shows the results of the computational screening calculations described in Example 13. Column 1 lists the heavy chain variable positions for the A chain, i.e. for only one of the C $\gamma$ 2-C $\gamma$ 3 heavy chains of the homodimer. Column 2 lists the wild type amino acid identity at each variable position. The remaining 20 columns indicate which of the 20 natural amino acids are favorable substitutions for each variable position according to the computational screening calculations. The presence of an amino acid in its column for a variable position indicates that the amino acid is within 1 unit of energy of the lowest energy substitution. Figure 31b shows an experimental library derived from the computational screening results, as described in Example 13. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The library is represented combinatorially, i.e. the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the library, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

25

Figure 32. rhumAb VEGF affinity maturation. The large central figure shows the 1CZ8 rhumAb VEGF V<sub>H</sub> and V<sub>L</sub> domains as gray ribbons bound to the VEGF target antigen as black ribbon, with Example 14 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

30

Figures 33a and 33b. rhumAb VEGF affinity maturation. Figure 33a shows the results of the computational screening calculations described in Example 14. Column 1 lists the light (L) and heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core, Surface, and Boundary classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT rhumAb VEGF amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figure 33b shows an experimental library derived from the computational screening results, as described in Example 14. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The libraries are represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all

40

5 other possible amino acid substitutions at all other positions. The complexity of the libraries, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

Figure 34. rhumAb VEGF affinity maturation. The large central figure shows the 1CZ8 rhumAb VEGF V<sub>H</sub> and V<sub>L</sub> domains as gray ribbons bound to the VEGF target antigen shown as black ribbon, with Example 14 variable position residues represented as black lines. The smaller figure in the upper left

10 shows the modeled full-length antibody structure with the relevant region highlighted by a box.

Figures 35a and 35b. rhumAb VEGF affinity maturation. Figure 35a shows the results of the computational screening calculations described in Example 14. Column 1 lists the light (L) and heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position.

15 The set of amino acids belonging to the Core, Surface, and Boundary classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT rhumAb VEGF amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are

20 observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figure 35b shows an experimental library derived from the computational screening results, as described in Example 14. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The libraries are represented combinatorially, that is the explicit

25 library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the libraries, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

Figure 36. SM3 affinity maturation. The large central figure shows the 1SM3 V<sub>H</sub> and V<sub>L</sub> domains as gray ribbons bound to the MUC1 antigen shown as black ribbon, with Example 15 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

30

Figures 37a, 37b, and 37c. SM3 affinity maturation. Figures 37a and 37b show the results of the computational screening calculations described in Example 15. Column 1 lists the light (L) and heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position.

35 The set of amino acids belonging to the Core, Surface, and Boundary classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT SM3 amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figure 37c shows an experimental library derived from the computational screening results, as described in Example 15.

40

5 Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The libraries are represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the libraries, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

10

Figure 38. Campath affinity maturation. The large central figure shows the 1CE1 V<sub>H</sub> and V<sub>L</sub> domains as gray ribbons bound to the CD52 antigen shown as black ribbon, with Example 16 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

15

Figures 39a and 37b. Campath affinity maturation. Figure 39a shows the results of the computational screening calculations described in Example 16. Column 1 lists the light (L) and heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core, Surface, and Boundary classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT Campath amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figure 39b shows an experimental library derived from the computational screening results, as described in Example 16. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The libraries are represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the libraries, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

Figures 40a and 40b. Sequence alignment of Campath variable region with the human variable region germ line. The Campath V<sub>H</sub> and V<sub>L</sub> sequences are shown aligned with the sequences that encode the human V<sub>H</sub> (Figure 40a) and V<sub>L</sub> (Figure 40b) germ line. The germ line sequences were obtained from the IMGT database, and numbered according to the numbering scheme of Chothia. The regions of the variable region are indicated above the numbering, and these include framework regions 1 through 3 (FR1, FR2, and FR3) and the complementarity determining regions (CDRs) 1 through 3 (CDR1, CDR2, and CDR3). Positions that make up CDRs are underlined. The 7 germ line subfamilies for V<sub>H</sub> and V<sub>L</sub> are grouped together and separated by a blank line. The Campath V<sub>H</sub> and V<sub>L</sub> sequences were aligned to the germ line sequences using the alignment program BLAST. Campath V<sub>H</sub> is most similar to the germ line chain VH\_4-59 and VH\_3-72, and Campath V<sub>L</sub> is most similar to the germ line chain VLk\_1D-33. The Campath V<sub>H</sub> and V<sub>L</sub> sequences are indicated by the underlined pdb accession code 1CE1, and shown below the subfamily to which they are closest in

5 sequence. Amino acids at variable positions for Example 16 design calculations are shown in bold in the 1CE1 and the germ line sequences.

10 Figures 41a and 41b. Campath sequence-guided affinity maturation. Figure 41a shows the results of the computational screening calculations described in Example 16. Rows 1 through 3 list the light (L) or heavy (H) chain variable positions, as defined in the 1CE1 structure, and the according to the Chothia numbering scheme. Row 4 lists the amino acids considered at variable positions as obtained from Figures 40a and 40b, and row 5 lists the amino acid at each position in the WT Campath sequence. "All" or "All 20" means that all 20 amino acids are considered at the variable position. The rows that follow list the amino acid identity at variable positions for the lowest energy sequence from 15 each cluster group, as described in Example 16. Figure 41b is similar to Figure 41a except that all the listed sequences are the set of sequences make up cluster groups 4 and 9.

20 Figure 42. D3H44 affinity maturation. The large central figure shows the 1JPS V<sub>H</sub> and V<sub>L</sub> domains as gray ribbons bound to the tissue factor antigen shown as black ribbon, with Example 16 variable position residues represented as black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

25 Figures 43a, 43b, 43c, and 43d. D3H44 affinity maturation. Figures 43a and 43b show the results of the computational screening calculations using the 1JPS template and 1JPT template respectively, as described in Example 17. Column 1 lists the light (L) and heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core, Surface, and Boundary classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT D3H44 amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground 30 state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figures 43c and 43d show an experimental library derived from the computational screening results, as described in Example 17. In Figure 43c, column 1 lists 35 variable positions, and columns 2 and 3 show amino acid substitutions, which are included in the experimental library. In Figure 43d, column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The libraries are represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The 40 complexity of the libraries, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

Figure 44. Herceptin affinity maturation. The large central figure shows the 1FVC V<sub>H</sub> and V<sub>L</sub> domains as black and gray ribbons respectively, with Example 18 variable position residues represented as

5 black lines. The smaller figure in the upper left shows the modeled full-length antibody structure with the relevant region highlighted by a box.

Figures 45a and 45b. Herceptin affinity maturation. Figure 45a shows the results of the computational screening calculations described in Example 18. Column 1 lists the light (L) and heavy (H) chain variable positions. Column 2 lists the amino acids considered at each variable position. The set of amino acids belonging to the Core, Surface, and Boundary classifications are described in the section entitled "Selection of Amino Acids to be Considered at Each Position". Column 3 lists the WT Herceptin amino acid identity at each variable position. Column 4 lists the amino acid identity at each variable position in the DEE ground state sequence predicted by the computational screening calculations. Column 5 lists the set of amino acids at each variable position that are observed in the Monte Carlo output. Each amino acid is followed by its occupancy, that is the number of sequences in the 1000 sequence set that contain that amino acid at that variable position. Figure 45b shows an experimental library derived from the computational screening results, as described in Example 18. Column 1 lists variable positions, and column 2 shows amino acid substitutions that are included in the experimental library. The libraries are represented combinatorially, that is the explicit library is the combination of each possible amino acid substitution at each variable position with all other possible amino acid substitutions at all other positions. The complexity of the libraries, that is the total number of defined sequences of which it is composed, is shown in the bottom row.

## 25 DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to the use of a variety of computational methods to alter physico-chemical properties of antibodies, to allow the virtual screening of large numbers of potential variants to arrive at sets that exhibit desirable properties as compared to the starting antibody or antibodies. The computational analyses can be done as a single step, with the resulting set being experimentally generated and tested in the desired assay, for improved function and properties. Similarly, the original set can be additionally computationally manipulated to create a new library which then itself can be experimentally tested.

35 The invention finds use in the prescreening of variant antibody libraries; that is, computational screening for stability (or other properties) may be done on either the entire protein or some subset of residues, as desired and described below. By using computational methods to generate a threshold or cutoff to eliminate disfavored sequences, the percentage of useful variants in a given variant set size can increase, and the required experimental outlay is decreased.

40 In order that the invention may be more completely understood, several definitions are set forth below. By "affinity maturation" herein is meant the process of enhancing the affinity of an antibody for its antigen. Methods for affinity maturation include but are not limited to computational screening methods and experimental methods. By "antibody" herein is meant a protein consisting of one or

5 more polypeptides substantially encoded (defined below) by all or part of the recognized antibody genes. The recognized immunoglobulin genes include, but are not limited to, the kappa, lambda, alpha, gamma (IgG1, IgG2, IgG3, and IgG4), delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Antibody herein is meant to include full-length antibodies and antibody fragments, and include antibodies that exist naturally in any organism or are

10 engineered (e.g. are variants). By "antibody fragment" is meant any form of an antibody other than the full-length form. Antibody fragments herein include antibodies that are smaller components that exist within full-length antibodies, and antibodies that have been engineered. Antibody fragments include but are not limited to Fv, Fc, Fab, and (Fab')<sub>2</sub>, single chain Fv (scFv), diabodies, triabodies, tetrabodies, bifunctional hybrid antibodies, and the like (Maynard & Georgiou, 2000, *Annu. Rev. Biomed. Eng.* 2:339-76; Hudson, 1998, *Curr. Opin. Biotechnol.* 9:395-402). By "amino acid" and "amino acid identity" as used herein is meant one of the 20 naturally occurring or any non-natural analogues that may be present at a specific, defined position. By "computational screening method" herein is meant any method for designing one or more mutations in a protein, wherein said method utilizes a computer to evaluate the energies of the interactions of potential amino acid side chain

20 substitutions with each other and/or with the rest of the protein. By "experimental library" herein is meant a list of one or more protein variants, existing either as a list of amino acid sequences or a list of the nucleotides sequences encoding them. Description of an experimental library may be defined, meaning that variant sequences are expressly described. Description of an experimental library may also be combinatorial, meaning that possible amino acid identities are indicated at variable positions, and the combination of all possibilities at all variable positions results in an expanded, explicitly

25 defined library. By "Fc" herein is meant the polypeptides of an antibody that are comprised of immunoglobulin domains Cgamma2 and Cgamma3 (Cγ2 and Cγ3). Fc may also include any residues which exist in the N-terminal hinge between Cγ2 and Cgamma1 (Cγ1). These regions are shown in Figure 1. Fc may refer to this region in isolation, or this region in the context of an antibody or

30 antibody fragment. By "full-length antibody" herein is meant the structure that constitutes the natural biological form of an antibody. In most mammals, including humans, and mice, this form is a tetramer and consists of two identical pairs of two immunoglobulin chains, each pair having one light and one heavy chain, each light chain comprising immunoglobulin domains V<sub>L</sub> and C<sub>L</sub>, and each heavy chain comprising immunoglobulin domains V<sub>H</sub>, Cγ1, Cγ2, and Cγ3. In each pair, the light and heavy chain variable regions (V<sub>L</sub> and V<sub>H</sub>) are together responsible for binding to an antigen, and the constant regions (C<sub>L</sub>, Cγ1, Cγ2, and Cγ3, particularly Cγ2, and Cγ3) are responsible for antibody effector functions. In some mammals, for example in camels and llamas, full-length antibodies may consist of only two heavy chains, each heavy chain comprising immunoglobulin domains V<sub>H</sub>, Cγ2, and Cγ3. By

35 "immunoglobulin (Ig)" herein is meant a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes. Immunoglobulins include but are not limited to antibodies. Immunoglobulins may have a number of structural forms, including but not limited to full-length antibodies, antibody fragments, and individual immunoglobulin domains including but not limited to V<sub>H</sub>, Cγ1, Cγ2, Cγ3, V<sub>L</sub>, and C<sub>L</sub>. By "immunoglobulin (Ig) domain" herein is meant a protein domain consisting of a polypeptide substantially encoded by an immunoglobulin gene. Ig domains include but

40

5 are not limited to  $V_H$ ,  $C\gamma 1$ ,  $C\gamma 2$ ,  $C\gamma 3$ ,  $V_L$ , and  $C_L$  as is shown in Figure 1. By "position" as used herein is meant a location in the sequence of a protein. Positions are typically, but not always, numbered sequentially. For example, position 297 is a position in the human antibody IgG1. By "residue" as used herein is meant a position in a protein and its associated amino acid identity. For example, Asparagine 297 (or Asn297 or N297) is a residue in the human antibody IgG1. By "variant protein  
10 sequence" as used herein is meant a protein sequence that has one or more residues that differ in amino acid identity from another similar protein sequence. Said similar protein sequence may be the natural wild type protein sequence, or another variant of the wild type sequence. In general, a starting sequence is referred to as a "parent" sequence, and again may either be a wild type or variant sequence. For example, preferred embodiments of the present invention may utilized humanized  
15 parent sequences upon which computational analyses are done. By "variable region" of an antibody herein is meant a polypeptide or polypeptides composed of the  $V_H$  immunoglobulin domain, the  $V_L$  immunoglobulin domains, or the  $V_H$  and  $V_L$  immunoglobulin domains as is shown in Figure 1 (including variants). Variable region may refer to this or these polypeptides in isolation, as an Fv fragment, as an scFv fragment, as this region in the context of a larger antibody fragment, or as this  
20 region in the context of a full-length antibody.

The present invention may be applied to antibodies obtained from a wide range of sources. The antibody may be substantially encoded by an antibody gene or antibody genes from any organism, including but not limited to humans, mice, rats, rabbits, camels, llamas, dromedaries, monkeys,  
25 particularly mammals and particularly human and particularly mice and rats. In a preferred embodiment, the antibody is fully human, obtained for example using transgenic mice or other animals (Bruggemann & Taussig, 1997, *Curr. Opin. Biotechnol.* 8:455-458) or human antibody libraries coupled with selection methods (Griffiths & Duncan, 1998, *Curr. Opin. Biotechnol.* 9:102-108). The antibody does not necessarily need to be naturally occurring. For example the present  
30 invention could be used to optimize an engineered antibody, including but not limited to chimeric antibodies and humanized antibodies (Clark, 2000, *Immunol. Today* 21:397-402). In addition, the antibody being optimized may be an engineered variant of an antibody that is substantially encoded by one or more natural antibody genes. For example, in a one embodiment the antibody being optimized is an antibody that has been affinity matured.

35 In general, the computationally generated antibody genes of the present invention are designed to be substantially encoded by a naturally occurring antibody gene such as a humanized antibody gene. "Substantially encoded" can include a number of components, including host cell codon usage and complementarity to wild type genes. For example, in one embodiment, "substantially encoded" can  
40 be defined as the ability of the computationally generated gene being sufficiently complementary to the wild type gene (or its complement, depending on sense and antisense considerations) such that hybridization can occur. This complementarity need not, and is preferably not perfect; that is, due to the alteration of the variable residues, there are a number of substitutions (and sometimes insertions or deletions) between the two sequences that result in differences between the sequences. However,

5 if the number of mutations is so great that no hybridization can occur under even the least stringent of hybridization conditions, the sequence is not a complementary sequence. Thus, by "substantially complementary" herein is meant that the sequences are sufficiently complementary to each other to hybridize under the selected reaction conditions. High stringency conditions are known in the art; see for example Maniatis et al., *Molecular Cloning: A Laboratory Manual*, 2d Edition, 1989, and Short  
10 *Protocols in Molecular Biology*, ed. Ausubel, et al., both of which are hereby incorporated by reference. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, "Overview of principles of hybridization and the  
15 strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10 C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength pH. The  $T_m$  is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at  $T_m$ , 50% of the probes are occupied at  
20 equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 C for short probes (e.g. 10 to 50 nucleotides) and at least about 60 C for long probes (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In another embodiment, less stringent  
25 hybridization conditions are used; for example, moderate or low stringency conditions may be used, as are known in the art; see Maniatis and Ausubel, *supra*, and Tijssen, *supra*.

In another embodiment, "substantially encoded" means that at least a significant portion of the gene is identical to the parent gene such as a humanized or human antibody. In preferred embodiments,  
30 there are large areas of perfect complementarity punctuated by the variant positions which may be different. In preferred embodiments, at least 75% of the total gene is encoded by the parent gene, with at least 85%, 90%, 95% and 98% being preferred.

The present invention may be applied to a wide range of antibody structural forms. For example, the  
35 antibody may be a full-length antibody, an antibody fragment, an Fc region, a variable region, an individual immunoglobulin domain, or a structural motif, site, or loop of an antibody. The antibody may comprise more than one protein chain. That is, the antibody may be an oligomer, including a homo- or hetero-oligomer.

40 The present invention may be applied to a wide range of antibody products. In one embodiment the antibody product is a therapeutic, a diagnostic, or a research reagent. In a preferred embodiment the antibody product is a therapeutic antibody which may be used to treat disease, such diseases including, but not limited to cancer, autoimmune disease, cardiovascular disease, and the like. The antibody product may find use in a composition that is monoclonal or polyclonal, and that could be

5 injected intravenously, subcutaneously, intramuscularly, and the like, as well as inhaled, applied topically, or via an oral dosage form, or otherwise administered. In an alternate embodiment, the antibody product is a library that could be screened experimentally, for example to generate antibodies against a target antigen using a selection method as described herein, or to affinity mature a particular antibody. This library may be a theoretical library, that is a list of nucleic acid or amino  
10 acid sequences, or may be a physical library of nucleic acids or proteins that encode the library sequences.

#### Computational Screening Methodology

A three-dimensional structure of an antibody is used as the starting point of the computational  
15 screening method of the present invention. The positions to be optimized are identified, which may be the entire antibody sequence or subset(s) thereof. Amino acids that will be considered at each position are selected. In a preferred embodiment, each considered amino acid may be represented by a discrete set of allowed conformations, called rotamers. Interaction energies are calculated between each considered amino acid and 1) each other considered amino acid, and 2) the rest of the  
20 protein, including the protein backbone and invariable residues. In a preferred embodiment, interaction energies are calculated between each considered amino acid side chain rotamer and 1) each other considered amino acid side chain rotamer and 2) the rest of the protein, including the protein backbone and invariable residues. One or more combinatorial search algorithms are then used to identify the lowest energy sequence and/or low energy sequences that will comprise an  
25 experimental library.

In a preferred embodiment, the computational screening method used to optimize antibodies is Protein Design Automation® (PDA™) technology, as is described in US 6,188,965; 6,269,312; and 6,403,312; USSNs 09/782,004; 09/927,790; and 10/218,102; PCTs 98/07254; 01/40091; and  
30 02/25588, all of which are expressly incorporated herein by reference. In another preferred embodiment, a Sequence Prediction Algorithm (SPA) is used to design proteins that are compatible with a known protein backbone structure as is described in Raha, *et al.*, 2000, *Protein Sci.* 9:1106-1119, USSNs 09/877,695 and 10/071,859, all expressly incorporated herein by reference. In some embodiments, combinations of different computational screening methods are used, including  
35 combinations of PDA™ and SPA, as well as combinations of these computational techniques in combination with sequence and structural alignment. Similarly, these computational methods can be used simultaneously or sequentially, in any order. Furthermore, these computational methods can be used with experimental methods (shuffling, error-prone PCR, etc.) as outlined below. It is also  
40 important to note that reiterative cycles are included; thus for example, a first computational step may be done, followed by some experimental techniques, followed by additional computational techniques.

Computational screening, viewed broadly, has four steps: 1) selection and preparation of the antibody template or templates, 2) selection of variable positions and considered amino acids at those positions, and in a preferred embodiment selection of rotamers to model amino acids, 3) energy

- 5 calculation, and 4) combinatorial optimization. As will be appreciated by those skilled in the art, energy calculation and combinatorial optimization are the computationally intensive aspects of computational screening, and together these two steps are referred to as design calculations.

#### Selection and Preparation of the Antibody Template

- 10 By "template antibody" herein is meant the structural coordinates of part or all of an antibody to be optimized. The template antibody is used as input in the computational screening calculations. A template protein may be part or all of any protein that has a known structure or for which a structure may be calculated, estimated, modeled or determined experimentally.
- 15 The template protein may be any antibody for which a three dimensional structure (that is, three dimensional coordinates for a set of the protein's atoms) is known or may be generated. The three dimensional structures of antibodies may be determined using methods including but not limited to X-ray crystallographic techniques, nuclear magnetic resonance (NMR) techniques, de novo modeling, and homology modeling. Antibody/antigen complexes may also be obtained using docking methods.
- 20 Suitable antibody structures include, but are not limited to, all of those found in the Protein Data Base compiled and serviced by the Research Collaboratory for Structural Bioinformatics (RCSB, formerly the Brookhaven National Lab).

- As will be appreciated by those skilled in the art, antibodies are a family of proteins that are closely
- 25 related in sequence and structure. Consequently, homology models, which are generated using available sequence and structure information from other antibodies, are often of high quality. Thus, if optimization is desired for an antibody for which the structure has not been solved experimentally, a suitable structural model may be generated that may serve as the template for design calculations. Methods for generating homology models are known in the art. Methods for generating homology
- 30 models of proteins are known in the art, and these methods find use in the present invention. See for example, Luo, *et al.* 2002, *Protein Sci.* 11: 1218-1226, Lehmann & Wyss, 2001, *Curr. Opin. Biotechnol.* 12(4):371-5.; Lehmann *et al.*, 2000, *Biochim Biophys Acta.* 1543(2):408-415; Rath & Davidson, 2000, *Protein Sci.*, 9(12):2457-69; Lehmann *et al.*, 2000, *Protein Eng.* 13(1):49-57; Desjarlais & Berg, 1993, *Proc Natl Acad Sci U S A.* 90(6):2256-60; Desjarlais & Berg, 1992, *Proteins.*
- 35 12(2):101-4; Henikoff & Henikoff, 2000, *Adv. Protein Chem.* 54:73-97; Henikoff & Henikoff, 1994, *J. Mol. Biol.* 243(4):574-8; all herein expressly incorporated by reference. Methods for generating homology models of antibodies in particular are described in Morea *et al.*, 2000, *Methods* 20:267-269, all herein expressly incorporated by reference.

- 40 As discussed above, the template may comprise any of a number of antibody structural forms. The template used in antibody design calculations may comprise an entire full-length antibody, a subset of an antibody such as a fragment, an individual immunoglobulin domain, or a structural motif, site, or loop of an antibody. The template antibody may comprise more than one protein chain, and may be the complex of an antibody bound to its antigen or to an antibody receptor. The template may

5 additionally contain nonprotein components, including but not limited to small molecules, substrates, cofactors, metals, water molecules, prosthetic groups, polymers and carbohydrates. As will be appreciated by those in the art, the target antigen of an antibody may be a protein or a non-protein molecule. In a preferred embodiment, the structural template is a plurality or set of template proteins, for example or an ensemble of structures such as those obtained from NMR. Alternatively, the set of  
10 antibody templates is generated from a set of related proteins or structures, or artificially created ensembles.

The protein template may be modified or altered prior to design calculations. A variety of methods for template preparation are described in US 6,188,965; 6,269,312; and 6,403,312; USSNs 09/782,004;  
15 09/927,790; 09/877,695; 10/071,859 and 10/218,102; PCTs 98/07254; 01/40091; and 02/25588, all of which are herein expressly incorporated by reference. For example, in a preferred embodiment, explicit hydrogens may be added if not included within the structure. In a preferred embodiment, energy minimization of the structure is run to relax strain, including strain due to van der Waals clashes, unfavorable bond angles, and unfavorable bond lengths. Alternatively, the protein template  
20 is altered using other methods, such as manually, including directed or random perturbations. It is also possible to modify the protein template during later steps of a design calculation, including during the energy calculation and combinatorial optimization steps. In an alternate embodiment, the protein template is not modified before or during design calculations.

## 25 Selection of Variable Positions and Considered Amino Acids

### *Selection of Variable, Floated, and Fixed Positions*

As is known in the art, it may be beneficial to reduce the complexity of a calculation by allowing mutation only at certain variable positions. By "variable position" herein is meant a position at which the amino acid identity is allowed to be altered in a design calculation. In a preferred embodiment the  
30 amino acid identity to which a position may be mutated is the full set or a subset of the 20 naturally occurring amino acids. Alternatively, variable positions may be allowed to mutate to a set of non-naturally occurring amino acids or synthetic analogs. One or more residues may be variable positions in design calculations.

35 Residues that are chosen as variable positions may be those that contribute to or are hypothesized to contribute to the antibody property to be optimized. For the present invention, these properties include stability, solubility, and affinity for antigen. Residues at variable positions may contribute favorably or unfavorably to a specific antibody property. For example, a residue at the antibody/antigen interface may be involved in mediating binding with antigen, and thus this position may be  
40 varied in design calculations aimed at improving affinity with antigen. Alternatively, as another example, a residue which has an exposed hydrophobic side chain may be responsible for causing unfavorable aggregation, and thus this position may be varied in design calculations aimed at improving solubility.

5 Thus in one embodiment, variable positions may be those positions that are directly involved in interactions that are determinants of an antibody property. For example, the antigen binding site of an antibody may be defined to include all residues that contact antigen. By "contact" herein is meant some chemical interaction between at least one atom of an antibody residue with at least one atom of the bound antigen, with chemical interaction including, but not limited to van der Waals interactions, 10 hydrogen bond interactions, electrostatic interactions, and hydrophobic interactions. In an alternative embodiment, variable positions may include those positions that are indirectly involved in an antibody property, i.e. such positions may be proximal to residues that contribute to an antibody property. For example, the antigen binding site of an antibody may be defined to include all residues within a certain distance, for example 4 - 10 Å, of the residues that are in van der Waals contact with antigen. Thus 15 variable positions in this case may be chosen not only as residues that directly contact antigen, but also those that contact residues that contact antigen and thus influence antigen binding indirectly. The specific positions chosen are dependent on the design strategy being employed.

20 In a preferred embodiment, some of the residue positions that are not variable are floated. By "floated position" herein is meant a position at which the amino acid conformation but not the amino acid identity is allowed to vary in a protein design calculation. In one embodiment the floated position may have the wild type amino acid identity. For example, floated positions may be wild type positions that are within a small distance of, for example, 5 Å, of a variable position residue. In an alternate embodiment, a floated position may have a non-wild type amino acid identity. Such an embodiment 25 may find use in the present invention, for example, when the goal is to evaluate the energetic or structural outcome of a specific mutation.

Residue positions that are not variable or floated are fixed. By "fixed position" herein is meant a position at which the amino acid identity and the conformation are held constant in a protein design calculation. Residues, which may be fixed, may include residues that are not involved or not thought 30 to be involved in the property to be optimized. In this case there is nothing to be gained by varying these positions. Residues that may be fixed may also include but are not limited to residues that are important for maintaining proper folding, structure, stability, solubility, and biological function. For example, residues that interact with protein receptors or residues that are glycosylation sites may be 35 fixed in design calculations to ensure that receptor binding and proper glycosylation respectively are not perturbed. Likewise, if stability is being optimized, it may be beneficial to fix residues that directly or indirectly interact with antigen so that antigen binding is not perturbed. Fixed positions may also include structurally important residues such as cysteines participating in disulfide bridges, residues critical for backbone conformation such as proline or glycine, critical hydrogen bonding residues, and 40 residues that form favorable packing interactions.

#### *Selection of Amino Acids to be Considered at Each Position*

The next step in the computational screening method of the present invention is to select a set of possible amino acid identities that will be considered at each particular variable position. This set of

5 possible amino acids is herein referred to as "considered amino acids" at a variable position. In one embodiment, all 20 amino acids (or their analogues or synthetic amino acids) are considered at a given variable position. Alternatively, a subset of amino acids, or even only one amino acid is considered at a given variable position. As will be appreciated by those skilled in the art, there is a computational benefit to considering only certain amino acid identities at variable positions, as it  
10 decreases the combinatorial complexity of the search. Furthermore, considering only certain amino acids at variable positions may be used to tune calculations toward specific design strategies. For example, for solubility optimization, it may be beneficial to allow only polar amino acids to be considered at surface exposed variable positions. In a preferred embodiment for solubility, at least one antibody sequence possesses an increase in polar character. Alternatively preferred, is selecting  
15 at least one nonpolar amino acid and substituting said nonpolar amino acid with a polar amino acid.

A wide variety of methods may be used, alone or in combination, to select which amino acids will be considered at each position, including but not limited to those discussed below.

20 For example, as is known in the art, the set of amino acids allowed at variable positions may be chosen based on the degree of exposure to solvent. Hydrophobic or nonpolar amino acids typically reside in the interior or core of a protein, which are inaccessible or nearly inaccessible to solvent. Thus at variable core positions it may be beneficial to consider only or mostly nonpolar amino acids such as alanine, valine, isoleucine, leucine, phenylalanine, tyrosine, tryptophan, and methionine.

25 Hydrophilic or polar amino acids typically reside on the exterior or surface of proteins, which have a significant degree of solvent accessibility. Thus at variable surface positions it may be beneficial to consider only or mostly polar amino acids such as alanine, serine, threonine, aspartic acid, asparagine, glutamine, glutamic acid, arginine, lysine and histidine. Some positions are partly exposed and partly buried, and are not clearly protein core or surface positions, in a sense serving as  
30 boundary residues between core and surface residues. Thus at such variable boundary positions it may be beneficial to consider both nonpolar and polar amino acids such as alanine, serine, threonine, aspartic acid, asparagine, glutamine, glutamic acid, arginine, lysine histidine, valine, isoleucine, leucine, phenylalanine, tyrosine, tryptophan, and methionine.

35 Determination of the degree of solvent exposure at variable positions may be by subjective evaluation or visual inspection of the antibody template by one skilled in the art of protein structural biology, or by the use of a variety of algorithms that are known in the art. Selection of amino acid types to be considered at variable positions may be aided or determined wholly by computational methods, such as calculation of solvent accessible surface area, or using algorithms which assess the orientation of  
40 the C $\alpha$ -C $\beta$  vectors relative to a solvent accessible surface, as outlined in US 6,188,965; 6,269,312; and 6,403,312; USSNs 09/782,004; 09/927,790; and 10/218,102; PCTs 98/07254; 01/40091; and 02/25588, and expressly herein incorporated by reference. In an embodiment, each variable position may be classified explicitly as a core, surface, or boundary position.

5 In an alternate embodiment, selection of the set of amino acids allowed at variable positions may be hypothesis-driven. Hypotheses for which amino acid types should be considered at variable positions may be derived by a subjective evaluation or visual inspection of the antibody template by one skilled in the art of protein structural biology. For example, if it is suspected that a hydrogen bonding  
10 hydrogen bonds may be considered even if the position is in the core. Likewise, if it is suspected that a hydrophobic packing interaction may be favorable at a variable position, nonpolar residues that have the capacity to form favorable packing interactions may be considered even if the position is on the surface. Other examples of hypothesis-driven approaches may involve issues of backbone flexibility or protein fold. As is known in the art, certain residues, for example proline, glycine, and  
15 cysteine, play important roles in protein structure and stability. Glycine enables greater backbone flexibility than all other amino acids, proline constrains the backbone more than all other amino acids, and cysteines may form disulfide bonds. It may therefore be beneficial to include one or more of these amino acid types to achieve a desired goal. Alternatively, it may be beneficial to exclude one or more of these amino acid types from the list of considered amino acids.

20 In an alternate embodiment, subsets of amino acids may be chosen to maximize coverage. In this case, additional amino acids with properties similar to that in the antibody template may be considered at variable positions. For example, if the residue at a variable position in the antibody template is a large hydrophobic residue, the user may choose to include additional large hydrophobic  
25 amino acids at that position. Alternatively, subsets of amino acids may be chosen to maximize diversity. In this case, amino acids with properties dissimilar to those in the antibody template may be considered at variable positions. For example, if the residue at a variable position in the antibody template is a large hydrophobic residue, the user may choose to include only one large hydrophobic amino acid in combination with other amino acids that are small, polar, etc.

30

5     *Selection of Rotamers to Model Amino Acids*

As is known in the art, some computational screening methods require only the identity of considered amino acids to be determined during design calculations. That is, no information is required concerning the conformations or possible conformations of the amino acid side chains. As is also known in the art, and in a preferred embodiment, a set of discrete side chain conformations, called  
10     rotamers, can be considered for each amino acid. Thus, a set of rotamers will be considered at each variable and floated position. Rotamers may be obtained from published rotamer libraries (see for example, Lovel *et al.*, 2000, *Proteins: Structure Function and Genetics* 40:389-408; Dunbrack & Cohen, 1997, *Protein Science* 6:1661-1681; DeMaeyer *et al.*, 1997, *Folding and Design* 2:53-66; Tuffery *et al.*, 1991, *J. Biomol. Struct. Dyn.* 8:1267-1289, Ponder & Richards, 1987, *J. Mol. Biol.*  
15     193:775-791). As is known in the art, rotamer libraries may be backbone-independent or backbone-dependent. Rotamers may also be obtained from molecular mechanics or *ab initio* calculations, and using other methods. In a preferred embodiment, a flexible rotamer model is used (see Mendes *et al.*, 1999, *Proteins: Structure, Function, and Genetics* 37:530-543). Similarly, artificially generated rotamers may be used, or augment the set chosen for each amino acid and/or variable position. In a  
20     preferred embodiment, at least one conformation that is not low in energy is included in the list of rotamers. In an alternatively preferred embodiment, the rotamer of the variable position residue in the antibody template is included in the list of rotamers allowed for that variable position in the design calculation. In an alternative embodiment, only the identity of each amino acid considered at variable positions is provided, and no specific conformational states of each amino acid are used during  
25     design calculations. That is, use of rotamers is not essential for computational screening.

*Use of Experimental Information*

In one embodiment of the present invention, experimental information may be used to guide the choice of variable positions, and/or the choice of considered amino acids at variable positions. As is  
30     known in the art, mutagenesis experiments are often carried out to determine the role of certain residues in protein structure and function, for example, which protein residues play a role in determining stability, or which residues make up the antigen binding site of an antibody. Data obtained from such experiments are useful in the present invention.

35     For example, variable positions for affinity maturation calculation could involve varying all positions at which mutation has been shown to affect binding. Similarly, the results from such an experiment may be used to guide the choice of allowed amino acid types at variable positions. For example, if certain types of amino acid substitutions are found to be favorable, sets, subsets, and/or similar types of those amino acids may be chosen to maximize coverage. In one embodiment, additional amino acids  
40     with properties similar to that or those that were found to be favorable experimentally may be considered at variable positions. For example, if experimental mutation of a variable position residue at the antigen interface to a large hydrophobic residue was found to be favorable, the user may choose to include additional large hydrophobic amino acids at that position in the computational screen.

5

As is known in the art, display and other selection technologies may be coupled with random mutagenesis to generate a list or lists of amino acid substitutions that are favorable for the selected property. Such a list or lists obtained from such experimental work find use in the present invention. For example, positions that are found to be invariable in such an experiment may be excluded as  
10 variable positions in computational screening calculations, whereas positions that are found to be more acceptable to mutation or respond favorably to mutation may be chosen as variable positions. Similarly, the results from such experiments may be used to guide the choice of allowed amino acid types at variable positions. For example, if certain types of amino acids arise more frequently in an experimental selection, subsets or similar types of those amino acids may be chosen to maximize  
15 coverage. In one embodiment, additional amino acids with properties similar to those that were found to be favorable experimentally may be considered at variable positions. For example, if selected mutations at a variable position that resides at the antigen interface are found to be uncharged polar amino acids, the user may choose to include additional uncharged polar amino acids, or perhaps charged polar amino acids, at that position.

20

#### *Use of Sequence Information*

In one embodiment of the present invention, sequence information may be used to guide choice of variable positions, and/or the choice of amino acids considered at variable positions. As is known in the art, all antibodies share a common structural scaffold and are homologous in sequence.  
25 Furthermore, there is a large amount of sequence and structural information available for the antibody family of proteins. These favorable aspects of antibodies may be used to gain insight into particular positions in the antibody family. As is known in the art, sequence alignments are often carried out to determine which antibody residues are conserved and which are not conserved. That is to say, by comparing and contrasting alignments of antibody sequences, the degree of variability at a position  
30 may be observed, and the types of amino acids that occur naturally at positions may be observed. Data obtained from such analyses are useful in the present invention.

The benefit of using sequence information to choose variable positions and considered amino acids at variable positions are several fold. For choice of variable positions, the primary advantage of using  
35 sequence information is that insight may be gained into which positions are more tolerant and which are less tolerant to mutation. Thus sequence information may aid in ensuring that quality diversity, i.e. mutations that are not deleterious to protein structure, stability, etc., is sampled computationally. The same advantage applies to use of sequence information to select amino acid types considered at variable positions. That is, the set of amino acids which occur in an antibody sequence alignment  
40 may be thought of as being pre-screened by evolution to have a higher chance than random for being compatible with an antibody's structure, stability, solubility, function, etc. Thus higher quality diversity is sampled computationally. A second benefit of using sequence information to select amino acid types considered at variable positions is that certain alignments may represent sequences that may be less immunogenic than random sequences. For example, if the amino acids considered at a given

5 variable position are the set of amino acids which occur at that position in an alignment of human germ line antibody sequences, those amino acids may be thought of as being pre-screened by nature for generating no or low immune response if the optimized antibody is used as a human therapeutic.

The source of the sequences may vary widely, and include one or more of the known databases, including but not limited to the Kabat database ([immuno.bme.nwu.edu](http://immuno.bme.nwu.edu); Johnson & Wu, 2001, *Nucleic Acids Res.* 29:205-206; Johnson & Wu, 2000, *Nucleic Acids Res.* 28:214-218), the IMGT database (IMGT, the international ImMunoGeneTics information system®; [imgt.cines.fr](http://imgt.cines.fr); Lefranc *et al.*, 1999, *Nucleic Acids Res.* 27:209-212; Ruiz *et al.*, 2000 *Nucleic Acids Res.* 28:219-221; Lefranc *et al.*, 2001, *Nucleic Acids Res.* 29:207-209; Lefranc *et al.*, 2003, *Nucleic Acids Res.* 31:307-310), and VBASE ([mrc-cpe.cam.ac.uk/vbase-ok.php?menu=901](http://mrc-cpe.cam.ac.uk/vbase-ok.php?menu=901)). Antibody sequence information can be obtained, compiled, and/or generated from sequence alignments of germ line sequences or sequences of naturally occurring antibodies from any organism, including but not limited to mammals. For example, Figures 2a and 2b list the aligned human V<sub>H</sub> and V<sub>L</sub> kappa germ line sequences, along with several antibody variable region sequences relevant to the examples of the present invention. Alternatively, antibody sequence information can be obtained from a database that is compiled privately. Other databases which are more general nucleic acid or protein databases, i.e. not particular to antibodies, for example including but are not limited to SwissProt ([expasy.ch/sprot/](http://expasy.ch/sprot/)), GenBank ([ncbi.nlm.nih.gov/Genbank](http://ncbi.nlm.nih.gov/Genbank)) and Entrez ([ncbi.nlm.nih.gov/Entrez/](http://ncbi.nlm.nih.gov/Entrez/)), and EMBL Nucleotide Sequence Database ([ebi.ac.uk/embl/](http://ebi.ac.uk/embl/)), may find use in the present invention. There are numerous sequence-based alignment programs and methods known in the art, and all of these find use in the present invention for generation of antibody sequence alignments.

Once alignments are made, sequence information can be used to guide choice of variable positions. Such sequence information can relate the variability, natural or otherwise, of a given position. Variability herein should be distinguished from variable position. By "variability" herein is meant the degree to which a given position in a sequence alignment shows variation in the types of amino acids that occur there. Variable position, to reiterate, is a position chosen by the user to vary in amino acid identity during a computational screening calculation. Variability may be determined qualitatively by one skilled in the art of bioinformatics. There are also methods known in the art to quantitatively determine variability that may find use in the present invention. The most preferred embodiment measures Information Entropy or Shannon Entropy. Variable positions can be chosen based on sequence information obtained from closely related antibody sequences, or antibody sequences that are less closely related.

40 The use of sequence information to choose variable positions finds broad use in the present invention. For example, to optimize antibody solubility by replacing exposed nonpolar surface residues, variable positions may be chosen as only that set of surface exposed positions that show a certain level of variability. As another example, to optimize antibody stability by mutating interdomain interface residues, variable positions may be chosen as only that set of interface positions that shown

5 a certain level of variability. For example, if an interface position in the antibody template is tryptophan, and tryptophan is observed at that position in greater than 90% of the sequences in an alignment, it may be beneficial to leave that position fixed. In contrast, if another interface position is found to have a greater level of variability, for example if five different amino acids are observed at that position with frequencies of approximately 20% each, that position may be chosen as a variable

10 position. In another embodiment, variable positions for affinity maturation calculations could be chosen to be all positions or a subset of positions which are determined by sequence alignment to make up a complementarity determining region (CDR) loop. Alternatively, variable positions could be chosen to be those residues that are determined by sequence alignment to contact a CDR loop. Thus, visual inspection of an aligned antibody sequence may substitute for visual inspection of an

15 antibody structure. This is due to the high level of both sequence and structural similarity in the antibody family. The rationale here is that those positions which typically contact a CDR in most antibody structures, for example, are hypothesized to be positions which contact a CDR in the antibody template being optimized in the calculation.

20 Sequence information can also be used to guide the choice of amino acids considered at variable positions. Such sequence information can relate to how frequently an amino acid, amino acids, or amino acid types (for example polar or nonpolar, charged or uncharged) occur, naturally or otherwise, at a given position. In one embodiment, the set of amino acids considered at a variable position in design calculations may comprise the set of amino acids that is observed at that position in the

25 alignment. Thus, the position-specific alignment information is used directly to generate the list of considered amino acids at a variable position in a computational screening calculation. Such a strategy is well known in the art. See for example Lehmann & Wyss, 2001, *Curr. Opin. Biotechnol.* 12(4):371-5.; Lehmann *et al.*, 2000, *Biochim Biophys Acta.* 1543(2):408-415; Rath & Davidson, 2000, *Protein Sci.*, 9(12):2457-69; Lehmann *et al.*, 2000, *Protein Eng.* 13(1):49-57; Desjarlais & Berg, 1993, *Proc Natl Acad Sci U S A.* 90(6):2256-60; Desjarlais & Berg, 1992, *Proteins.* 12(2):101-4; Henikoff & Henikoff, 2000, *Adv. Protein Chem.* 54:73-97; Henikoff & Henikoff, 1994, *J. Mol. Biol.* 243(4):574-8; all

30 herein expressly incorporated by reference.

In an alternate embodiment, the set of amino acids considered at a variable position or positions may

35 comprise a set of amino acids that is observed most frequently in the alignment. Thus, a certain criteria is applied to determine whether the frequency of an amino acid or amino acid type will be included in the set of amino acids that are considered at a variable position in a design calculation. As is known in the art, sequence alignments may be analyzed using statistical methods to calculate the sequence diversity at any position in the alignment and the occurrence frequency or probability of

40 each amino acid at a position. Such data may then be used to determine which amino acids types to consider. In the simplest embodiment, these occurrence frequencies are calculated by counting the number of times an amino acid is observed at an alignment position, then dividing by the total number of sequences in the alignment. In other embodiments, the contribution of each sequence, position or amino acid to the counting procedure is weighted by a variety of possible mechanisms. In a preferred

5 embodiment, the contribution of each aligned sequence to the frequency statistics is weighted according to its diversity weighting relative to other sequences in the alignment. A common strategy for accomplishing this is the sequence weighting system recommended by Henikoff and Henikoff (see Henikoff & Henikoff, 2000, *Adv. Protein Chem.* 54:73-97; Henikoff & Henikoff, 1994, *J. Mol. Biol.* 243:574-8; both herein expressly incorporated by reference. In a preferred embodiment, the  
10 contribution of each sequence to the statistics is dependent on its extent of similarity to the target sequence, i.e. the antibody template used in the design calculations, such that sequences with higher similarity to the target sequence are weighted more highly. Examples of similarity measures include, but are not limited to, sequence identity, BLOSUM similarity score, PAM matrix similarity score, and Blast score. In an alternate embodiment, the contribution of each sequence to the statistics is  
15 dependent on its known physical or functional properties. These properties include, but are not limited to, thermal and chemical stability, contribution to activity, solubility, etc. For example, when optimizing an antibody for solubility, those sequences in an alignment that are known to be most soluble (for example see Ewert *et al.*, 2003, *J. Mol. Biol.* 325:531-553), will contribute more heavily to the calculated frequencies.

20  
Regardless of what criteria are applied for choosing the set of amino acids in a sequence alignment to be considered at variable positions, using sequence information to choose the set of amino acids considered at variable positions finds broad use in the present invention. For example, to optimize antibody solubility by replacing exposed nonpolar surface residues, considered amino acids may be  
25 chosen as the set of amino acids, or a subset of those amino acids which meet some criteria, that are observed at that position in an alignment of antibody sequences. As another example, to optimize antibody stability by mutating domain interface residues, considered amino acids may be chosen as the set of amino acids, or a subset of those amino acids that meet some criteria, that are observed at that position in an alignment of antibody sequences. In an alternate embodiment, one or more amino  
30 acids may be added or subtracted subjectively from a list of amino acids derived from a sequence alignment in order to maximize coverage. For example, additional amino acids with properties similar to those that are found in a sequence alignment may be considered at variable positions. For example, if an antigen binding position is observed to have uncharged polar amino acids in an antibody sequence alignment, the user may choose to include additional uncharged polar amino acids  
35 in an affinity maturation calculation, or perhaps charged polar amino acids, at that position.

In a preferred embodiment, sequence alignment is not used alone in the analysis step of the present invention; that is, sequence information is combined with energy calculation, as discussed below. For example, pseudo energies can be derived from sequence information to generate a scoring function.  
40 The use of a sequence-based scoring function may assist in significantly reducing the complexity of a calculation. However, as is appreciated by those skilled in the art, the use of a sequence-based scoring function alone may be inadequate because sequence information can often indicate misleading correlations between mutations that may in reality be structurally conflicting. Thus, in a preferred embodiment, a structure-based method of energy calculation is used, either alone or in

5 combination with a sequence-based scoring function. That is, preferred embodiments do not rely on sequence alignment information alone as the analysis step.

#### Energy Calculation

10 Some method of scoring each amino acid substitution, herein referred to as energy calculation, is required for computational screening. As previously discussed, there are a variety of ways to represent amino acids in order to enable efficient energy calculation.

15 In a preferred embodiment, considered amino acids are represented as rotamers, as described previously, and the energy (or score) of interaction of each possible rotamer at each variable position, or at each variable and floated position, with the template and/or other rotamers, is calculated. It should be understood that the template in this case includes both the atoms of the protein structure backbone, as well as the atoms of any fixed residues, as well as non-protein atoms. In a preferred embodiment, two sets of interaction energies are calculated for each side chain rotamer at every position: the interaction energy between the rotamer and the template (the "singles" energy), and the  
20 interaction energy between the rotamer and all other possible rotamers at every other variable and floated position (the "doubles" energy). In an alternate embodiment, singles and doubles energies are calculated for fixed positions as well as for variable and floated positions.

25 In an alternate embodiment, considered amino acids are not represented as rotamers.

In one embodiment, molecular dynamics calculations may be used to computationally screen sequences by individually calculating mutant sequence scores.

30 Regardless of how amino acids are represented, the energies of interaction are measured by one or more scoring functions. A variety of scoring functions find use in the present invention for calculating energies. As will be appreciated by those skilled in the art, certain scoring functions are more compatible with certain types of methods for representing amino acids. For example, force fields are particularly well suited to score amino acid substitutions that are represented as rotamers. However, in order to not constrain the present invention to any particular application or theory of operation, a  
35 variety of scoring functions are presented that may find use in the present invention regardless of how amino acids are represented.

40 Scoring functions may include a number of potentials, herein referred to as the energy terms of a scoring function, including but are not limited to, a van der Waals potential scoring function, a hydrogen bond potential scoring function, an atomic solvation potential scoring function, a secondary structure propensity potential scoring function and an electrostatic potential scoring function. At least one energy term is used to score each variable or floated position, although the energy terms may differ depending on the position classification or other considerations.

5 A variety of scoring functions are described in US 6,188,965; 6,269,312; and 6,403,312; USSNs  
09/782,004; 09/927,790; 09/877,695; 10/071,859 and 10/218,102; PCTs 98/07254; 01/40091; and  
02/25588, all of which are herein expressly incorporated by reference. As will be appreciated by  
those skilled in the art, a number of force fields, which are comprised of one or more energy terms,  
may serve as scoring functions. Force fields include, but are not limited to, *ab initio* or quantum  
10 mechanical force fields, semi-empirical force fields, and molecular mechanics force fields. In an  
alternate embodiment, scoring functions that are knowledge-based may be used. In an alternate  
embodiment, scoring functions that use statistical methods may find use in the present invention.  
These methods may be used to assess the match between a sequence and a three-dimensional  
protein structure, and hence may be used to score amino acid substitutions for fidelity to the protein  
15 structure.

In a preferred embodiment, additional energy terms may be included in the scoring function. For  
example, the above mentioned scoring functions may be modified to include terms including but not  
limited to torsional potentials, entropy potentials, additional solvation models including contact models,  
20 solvent exclusion models, and knowledge-based energies derived from protein sequence and/or  
structure statistics including but not limited to threading potentials, reference energies, pseudo  
energies, and sequence biases derived from sequence alignments (as discussed in the previous  
section). In a preferred embodiment, a scoring function is modified to include models for  
immunogenicity, such as functions derived from data on binding of peptides to MHC (Major  
25 Histocompatibility Complex), that may be used to identify potentially immunogenic sequences (see  
USSNs 09/903,378; 10/039,170; 60/222,697 and USSN to be determined, filed January 8, 2003 and  
entitled "NOVEL PROTEIN WITH ALTERED IMMUNOGENICITY"; and PCT 01/21823; and 02/00165,  
all herein expressly incorporated by reference).

30 In one embodiment, as is known in the art, one or more scoring functions may be optimized or  
"trained" during the computational analysis, and then the analysis re-run using the optimized system.  
Such altered scoring functions may be obtained for example, by training a scoring function using  
experimental data.

35 In a preferred embodiment, the scoring functions used are one or more of the scoring functions which  
are described in US 6,188,965; 6,269,312; and 6,403,312; USSNs 09/782,004; 09/927,790;  
09/877,695; 10/071,859 and 10/218,102; PCTs 98/07254; 01/40091; and 02/25588, all herein  
expressly incorporated by reference. In an alternate embodiment, energy calculation is carried out  
using one or more of the methods described above in combination.

40 In the most preferred embodiment, a scoring function using more than one energy term is used. As  
will be appreciated by those skilled in the art, Ig domain stabilization using only a van der Waals  
potential (Looger & Hellinga, 2001, *J. Mol. Biol.* 307:429-445) or affinity maturation using only an  
electrostatic potential may be inadequate for accurately evaluating the complex interactions in an

antibody and between an antibody and its antigen. In the most preferred embodiment, energies may be calculated using a force field containing energy terms describing van der Waals, solvation, electrostatic, hydrogen bond interactions and combinations thereof. In additional embodiments, additional energy terms include but are not limited to entropic terms, torsional energies, and knowledge-based energies.

#### Combinatorial Optimization

An important component of computational screening is the identification of one or more sequences that have a favorable score or are low in energy. In a preferred embodiment, all possible interaction energies are calculated prior to optimization. In an alternatively preferred embodiment, energies may be calculated as needed during optimization.

The need for a combinatorial optimization algorithm is illustrated by examining the number of possibilities that are considered in a typical design calculation. The discrete nature of rotamer sets allows a simple calculation of the number of possible rotameric sequences for a given design problem. A backbone of length  $n$  with  $m$  possible rotamers per position will have  $m^n$  possible rotamer sequences, a number which grows exponentially with sequence length. For very simple design calculations, it is possible to examine each possible sequence in order to identify the optimal sequence and/or one or more favorable sequences. However, for a typical design problem, the number of possible sequences (up to  $10^{80}$  or more) is sufficiently large that examination of each possible sequence is intractable. A variety of combinatorial optimization algorithms may then be used to identify the optimum sequence and/or one or more favorable sequences.

Combinatorial optimization algorithms may be divided into two classes: (1) those that are guaranteed to return the global minimum energy configuration if they converge, and (2) those that are not guaranteed to return the global minimum energy configuration, but which will always return a solution. Examples of the first class of algorithms include, but are not limited to, Dead-End Elimination (DEE) and Branch & Bound (B&B) (including Branch and Terminate) (Gordon & Mayo, 1999, *Structure Fold. Des.* 7:1089-98). Examples of the second class of algorithms include, but are not limited to, Monte Carlo (MC), self-consistent mean field (SCMF), Boltzmann sampling (Metropolis *et al.*, 1953, *J. Chem. Phys.* 21:1087), simulated annealing (Kirkpatrick *et al.*, 1983, *Science*, 220:671-680), genetic algorithm (GA) and Fast and Accurate Side-Chain Topology and Energy Refinement (FASTER (Desmet, *et al.*, 2002, *Proteins*, 48:31-43). A combinatorial optimization algorithm may be used alone or in conjunction with another combinatorial optimization algorithm.

In one embodiment of the present invention, the strategy for applying a combinatorial optimization algorithm is to find the global minimum energy configuration. In an alternate embodiment, the strategy is to find one or more low energy or favorable sequences. In an alternate embodiment, the strategy is to find the global minimum energy configuration and then find one or more low energy or favorable sequences. For example, as outlined in US 6,269,312 and PCT US98/07254, preferred

embodiments utilize a Dead End Elimination (DEE) step, and preferably a Monte Carlo step. In other embodiments tabu search algorithms are used or combined with DEE and/or Monte Carlo, among other search methods (see *Modern Heuristic Search Methods*, edited by V.J. Rayward-Smith, *et al.*, 1996, John Wiley & Sons Ltd., hereby expressly incorporated by reference in its entirety and also USSN 10/218,102 and PCT 02/25588). In another preferred embodiment, a genetic algorithm may be used. See, USSN 09/877,695 and 10/071,859, both herein expressly incorporated by reference. As another example, as is more fully described in US 6,188,965; 6,269,312; and 6,403,312; USSNs 09/782,004; 09/927,790; and 10/218,102; PCTs 98/07254; 01/40091; and 02/25588, which are herein expressly incorporated by reference, the global optimum may be reached, and then further computational processing may occur, which generates additional optimized sequences.

In the simplest embodiment, design calculations are not combinatorial. That is, energy calculations are used to evaluate the amino acid substitutions individually at single variable positions. However, it is a more preferred embodiment in certain situations to combine design calculations and also to evaluate amino acid substitutions at more than one variable positions.

#### Library Generation

The output sequence or sequences from computational screening may be used to generate an experimental library. By "experimental library" herein is meant a list of one or more protein variants, existing either as a list of amino acid sequences or a list of the nucleotides sequences encoding them.

Such a library may then be screened experimentally to single out superior members of antibody variants that are optimized for the desired property. As discussed above, computationally screened libraries have a number of benefits. Computationally generated libraries are significantly enriched in stable, properly folded, and functional sequences relative to randomly generated libraries. Because of the overlapping sequence constraints on antibody structure, stability, solubility, function, etc., a large number of the candidates in an experimental library occupy "wasted" sequence space. For example, a large fraction of sequence space encodes unfolded, misfolded, incompletely folded, partially folded, or aggregated proteins. In contrast, experimental libraries that are screened computationally are composed primarily of productive sequence space. As a result, computational screening increases the chances of identifying antibodies that are broadly optimized for stability, solubility, and affinity for antigen. In effect, computational screening yields an increased hit-rate, thereby decreasing the number of variants that must be screened experimentally. The term "experimental library" may refer to the set of optimized antibodies in any form. In one embodiment, the library is a list of nucleic acid or amino acid sequences, or a list of nucleic acid or amino acid substitutions at variable positions. For example, the examples used to illustrate the present invention below provide experimental libraries as amino acid substitutions at variable positions. In an alternate embodiment, the library is a physical library composed of nucleic acids that encode the optimized library sequences. Said nucleic acids may be the genes encoding the optimized antibodies, the genes encoding the optimized antibodies with any operably linked nucleic acids, or expression vectors encoding the library members together with any other operably linked regulatory sequences, selectable markers, fusion constructs, and/or

5 other elements. For example, the experimental library may be a set of mammalian expression vectors that encode library members, the protein products of which may be subsequently expressed, purified, and screened experimentally. As another example, the experimental library may be a display library. Such a library could, for example, be composed of a set of expression vectors which encode library members operably linked to some fusion partner that enables phage display, ribosome display, yeast  
10 display, bacterial surface display, and the like. Such a library could be used, for example, to screen for antibodies against a target antigen, or to affinity mature a particular antibody. In an alternate embodiment, the library is a physical library that is comprised of the optimized antibody proteins, either in purified or unpurified form.

15 In one embodiment, an experimental library is a list of at least one sequence that are variant antibodies optimized for a desired property. For example see, Filikov *et al.*, 2002, *Protein Sci.* 11:1452-1461 and Luo *et al.*, 2002, *Protein Sci* 11:1218-1226. In an alternate embodiment, an experimental library may be defined as a combinatorial list, meaning that each a list of amino acid substitutions is designed for each variable position, with the implication that each substitution is to be  
20 combined with all other designed substitutions at all other variable positions. In this case, expansion of the combination of all possibilities at all variable positions results in a large explicitly defined library.

#### *Selecting Sequences for the Experimental Library*

As is known in the art, there are a variety of ways that an experimental library may be derived from  
25 the output of computational screening calculations. For example, methods of library generation described in US 6,403,312; USSNs 09/782,004; 09/927,790; and 10/218,102; PCTs 01/40091; and 02/25588, herein expressly incorporated by reference, find use in the present invention.

In one embodiment, sequences scoring within a certain range of the global optimum sequence may  
30 be included in the library. For example, all sequences within 10 kcal/mol of the lowest energy sequence could be used as the experimental library. In an alternate embodiment, sequences scoring within a certain range of one or more local minima sequences may be used. In a preferred embodiment, the library sequences are obtained from a filtered set. Such a list or set may be generated by a variety of methods, as is known in the art, for example using an algorithm such as  
35 Monte Carlo, B&B, or SCMF. For example, the top  $10^3$  or the top  $10^5$  sequences in the filtered set may comprise the experimental library. Alternatively, the total number of sequences defined by the combination of all mutations may be used as a cutoff criterion for the experimental library. Preferred values for the total number of recombined sequences range from 10 to  $10^{20}$ , particularly preferred values range from 100 to  $10^9$ . Alternatively, a cutoff may be enforced when a predetermined number  
40 of mutations per position is reached.

Clustering algorithms may be useful for classifying sequences derived by computational screening methods into representative groups. For example, methods of clustering and their application described in USSN 10/218,102 and PCT 02/25588, herein expressly incorporated by reference, find

5 use in the present invention. Representative groups may be defined, for example, by similarity. Measures of similarity include, but are not limited to sequence similarity and energetic similarity. Thus the output sequences from computational screening may be clustered around local minima, referred to herein as clustered sets of sequences. For example, sets of sequences that are close in sequence space may be distinguished from other sets. In one embodiment, coverage within one or a subset of  
10 clustered sets may be maximized by including in the experimental library some, most, or all of the sequences that make up one or more clustered sets of sequences. For example, the user may wish to maximize coverage within the one, two, or three lowest energy clustered sets by including the majority of sequences within these sets in the library. In an alternate embodiment, diversity across clustered sets of sequences may be sampled by including within an experimental library only a subset  
15 of sequences within each clustered set. For example, all or most of the clustered sets could be broadly sampled by including the lowest energy sequence from each clustered set in the experimental library.

In some embodiments, sequences that do not make the cutoff are included in the experimental library.  
20 This may be desirable in some situations, for instance to evaluate the approach to library generation, to provide controls or comparisons, or to sample additional sequence space. For example, the WT antibody sequence may be included in the library, even if it does not make the cutoff.

The set of antibody sequences in an experimental library is generally, but not always, significantly  
25 different from the wild type antibody template, although in some cases the library preferably contains the wild-type sequence. The range of optimized protein sequences is dependent upon many factors including the size of the protein, properties desired, etc.

#### *Use of Sequence Information to Guide Library Generation*

30 In one embodiment of the present invention, sequence information may be used to guide or filter a computationally screened output for generation of an experimental library. As discussed, by comparing and contrasting alignments of antibody sequences, the degree of variability at a position and the types of amino acids which occur naturally at that position may be observed. Data obtained from such analyses are useful in the present invention. The benefits of using sequence information  
35 have been discussed, and those benefits apply equally to use of sequence information to guide library generation. The set of amino acids which occur in an antibody sequence alignment may be thought of as being pre-screened by evolution to have a higher chance than random at being compatible with an antibody's structure, stability, solubility, function, etc. Furthermore, certain alignments may provide represent sequences that are less immunogenic than random sequences. The variety of sequence  
40 sources, as well as the methods for generating antibody sequence alignments that have been discussed find use in the application of sequence information to guiding library generation. Likewise, as discussed above, various criteria may be applied to determine the importance or weight of certain residues in an alignment. These methods also find use in the application of sequence information to guide library generation.

5

Using sequence information to guide library generation from the results of computational screening finds broad use in the present invention. In one embodiment, sequence information is used to filter sequences from computational screening output. That is to say, some substitutions are subtracted from the computational output to generate the experimental library. For example, to optimize antibody solubility by replacing exposed nonpolar surface residues, the resulting output of a computational screening calculation or calculations may be filtered so that the experimental library includes only those amino acids, or a subset of those amino acids which meet some criteria, that are observed at that position in an alignment of antibody sequences. In an alternate embodiment, sequence information is used to add sequences to the computational screening output. That is to say, sequence information is used to guide the choice of additional amino acids that are added to the computational output to generate the experimental library. For example, to optimize antibody stability by mutating domain interface residues, the output set of amino acids for a given position from a computational screening calculation may be augmented to include one or more amino acids that are observed at that position in an alignment of antibody sequences. In an alternate embodiment, based on sequence alignment information, one or more amino acids may be added to or subtracted from the computational screening sequence output in order to maximize coverage or diversity. For example, additional amino acids with properties similar to those that are found in a sequence alignment may be added to the experimental library. For example, if a position involved in antigen binding is observed to have uncharged polar amino acids in an antibody sequence alignment, the user may choose to include additional uncharged polar amino acids to the experimental library at that position.

#### *Generation of Secondary Libraries*

In one embodiment of the present invention, libraries may be processed further to generate subsequent libraries. In this way, the output from a computational screening calculation or calculations may be thought of as a primary library. This primary library may be combined with other primary libraries from other calculations or other experimental libraries, processed using subsequent calculations, sequence information, or other analyses, or processed experimentally to generate a subsequent library, herein referred to as a secondary library, which could become an experimental library. As will be appreciated from this description, the use of sequence information to guide or filter libraries, discussed above, is itself one method of generating secondary libraries from primary libraries. Generation of secondary libraries gives the user greater control of the parameters within an experimental library. This enables more efficient experimental screening, and may allow feedback from experimental results to be interpreted more easily, providing a more efficient design/experimentation cycle.

40

There are a wide variety of methods to generate secondary libraries from primary libraries. For example, USSN 10/218,102 and PCT 02/25588, herein expressly incorporated by reference, describes methods for secondary library generation that find use in the present invention. Typically some selection step occurs in which a primary library is processed in some way. For example, in one

5 embodiment a selection step occurs where some set of primary sequences are chosen to form the secondary library. In an alternate embodiment, a selection step is a computational step, again generally including a selection step, wherein some subset of the primary library is chosen and then subjected to further computational analysis, including both further computational screening as well as techniques such as "in silico" shuffling (recombination). See, for example US 5,830,721; 5,811,238; 10 5,605,793; 5,837,458, PCT US/19256, Rachitt-Enchira (.enchira.com/gene\_shuffling.htm); error-prone PCR, for example using modified nucleotides; known mutagenesis techniques including the use of multi-cassettes; DNA shuffling (Crameri *et al.*, 1998, *Nature* 391:288-291); heterogeneous DNA samples (US 5,939,250); ITCHY (Ostermeier *et al.*, 1999, *Nat. Biotechnol.* 17:1205-1209); StEP (Zhao *et al.*, 1998, *Nat. Biotechnol.* 16:258-261), GSSM (US 6,171,820 and US 5,965,408); in vivo 15 homologous recombination, ligase assisted gene assembly, end-complementary PCR, profusion (Roberts & Szostak, 1997, *Proc. Natl. Acad. Sci. USA* 94:12297-12302); yeast/bacteria surface display (Lu *et al.*, 1995, *Biotechnology* 13:366-372); Seed & Aruffo, 1987, *Proc. Natl. Acad. Sci. USA* 84(10):3365-3369; Boder & Wittrup, 1997, *Nat. Biotechnol.* 15:553-557). all hereby incorporated by reference. In an alternate embodiment, a selection step occurs that is an experimental step, for 20 example any of the experimental library screening steps below, wherein some subset of the primary library is chosen and then recombined experimentally, for example using one of the directed evolution methods discussed below, to form a secondary library. In a preferred embodiment, the primary library is generated and processed as outlined in US 6,403,312, which is herein expressly incorporated by reference.

25 Generation of secondary and subsequent libraries finds broad use in the present invention. In one embodiment, different primary libraries may be combined to generate a secondary or subsequent library. In another embodiment, secondary libraries may be generated by sampling sequence diversity at highly mutable or highly conserved positions. The primary library may be analyzed to 30 determine which amino acid positions in the template protein have high mutational frequency, and which positions have low mutational frequency. For example, positions in an antibody that show a great deal of mutational diversity in computational screening may be fixed in a subsequent round of design calculations. A filtered set of the same size as the first would now show diversity at positions that were largely conserved in the first library. Alternatively, the secondary library may be generated 35 by varying the amino acids at the positions that have high numbers of mutations, while keeping constant the positions that do not have mutations above a certain frequency.

This discussion is not meant to constrain generation of libraries subsequent to primary libraries to secondary libraries. As will be appreciated, primary and secondary libraries may be processed further 40 to generate tertiary libraries, quaternary libraries, and so on. In this way, library generation is an iterative process. For example, tertiary libraries may be constructed using a variety of additional steps applied to one or more secondary libraries; for example, further computational processing may occur, secondary libraries may be recombined, or subsets of different secondary libraries may be combined. In a preferred embodiment, a tertiary library may be generated by combining secondary libraries. For

5 example, primary and/or secondary libraries that analyzed different parts of a protein may be combined to generate a tertiary library that treats the combined parts of the protein. In an alternate embodiment, the variants from a primary library may be combined with the variants from a second library to provide a combined tertiary library at lower computational cost than creating a very long filtered set. These combinations may be used, for example, to analyze large proteins, especially large  
10 multi-domain proteins. Thus the above description of secondary library generation applies to generating any library subsequent to a primary library, the end result being a final library that may be screened experimentally to obtain optimized antibodies. These examples are not meant to constrain generation of secondary libraries to any particular application or theory of operation for the present invention. Rather, these examples are meant to illustrate that generation of secondary libraries, and  
15 subsequent libraries such as tertiary libraries and so on, is broadly useful in computational screening methodology for experimental library generation.

### Experimental Library Screening

Once an experimental library is designed using any of the methods outlined herein or combinations thereof, the physical library may be constructed using a variety of techniques. The library may then  
20 be screened to obtain antibodies optimized for greater stability, solubility, and/or enhanced affinity for antigen. Accordingly, the present invention provides a variety of methods for constructing and screening experimental libraries. These methods are not meant to constrain the present invention to any particular application or theory of operation. Rather, the provided examples are meant to  
25 illustrate generally that computationally screened libraries may be screened experimentally to obtain antibodies with optimized physico-chemical properties. General methods for antibody molecular biology, expression, purification, and screening are described in Antibody Engineering, 2001, edited by Duebel & Kontermann, Springer-Verlag, Heidelberg; Hayhurst & Georgiou, 2001, *Curr. Opin. Chem. Biol.* 5:683-689; Maynard & Georgiou, 2000, *Annu. Rev. Biomed. Eng.* 2:339-76; all of which  
30 are herein expressly incorporated by reference.

### Molecular Biology and Library Generation

In one embodiment of the present invention, the experimental library sequences are used to create nucleic acids such as DNA which encode the antibody member sequences and which may then be  
35 cloned into host cells, expressed and assayed, if desired. Thus, nucleic acids, and particularly DNA, may be made which encode each member protein sequence. These practices are carried out using well-known procedures. For example, a variety of methods that may find use in the present invention are described in Molecular Cloning- A Laboratory Manual, 3<sup>rd</sup> Ed. (Maniatis, Cold Spring Harbor Laboratory Press, New York, 2001), and Current Protocols in Molecular Biology (Wiley & Sons,  
40 mrw2.interscience.wiley.com/cponline/), both of which are herein expressly incorporated by reference.

As will be appreciated by those in the art, the generation of exact sequences for a library comprising a large number of sequences is potentially expensive and time consuming. Accordingly, there are a variety of techniques that may be used to efficiently generate experimental libraries of the present

5 invention. Such methods that may find use in the present invention are described or referenced in US  
6,403,312; USSN 09/782,004; 09/927,790 and 10/218,102; and PCTs 01/40091 and 02/25588, all  
hereby incorporated by reference. Such methods include but are not limited to gene assembly  
methods, PCR-based method and methods which use variations of PCR, ligase chain reaction-based  
10 methods and methods which use oligos with random mutations, classical site-directed mutagenesis  
methods, cassette mutagenesis, and other amplification and gene synthesis methods. As is known in  
the art, there are a variety of commercially available kits and methods for gene assembly,  
mutagenesis, vector subcloning, and the like, and such commercial products find use in the present  
invention for generating nucleic acids that encode members of an experimental library.

#### Protein Expression

##### *Expression Systems*

The library antibody proteins of the present invention may be produced by culturing a host cell  
transformed with nucleic acid, preferably an expression vector, containing nucleic acid encoding an  
20 library protein, under the appropriate conditions to induce or cause expression of the library protein.  
The conditions appropriate for library protein expression will vary with the choice of the expression  
vector and the host cell, and will be easily ascertained by one skilled in the art through routine  
experimentation.

25 A wide variety of appropriate host cells may be used, including but not limited to mammalian cells,  
bacteria, insect cells, and yeast. For example, a variety of cell lines that may find use in the present  
invention are described in the ATCC cell line catalog (atcc.org), herein expressly incorporated by  
reference. .

30 In a preferred embodiment, the library proteins are expressed in mammalian expression systems,  
including systems in which the expression constructs are introduced into the mammalian cells using  
virus such as retrovirus or adenovirus. Any mammalian cells may be used, with mouse, rat, primate  
and human cells being particularly preferred. Suitable cells also include known research cells,  
including but not limited to Jurkat T cells, NIH3T3 cells, CHO, COS, etc. In an alternately preferred  
35 embodiment, library proteins are expressed in bacterial systems. Bacterial expression systems are  
well known in the art, and include *Escherichia coli* (*E. coli*), *Bacillus subtilis*, *Streptococcus cremoris*,  
and *Streptococcus lividans*. In an alternate embodiment, library proteins are produced in insect cells.  
In an alternate embodiment, library proteins are produced in yeast cells. In an alternate embodiment  
library proteins are expressed in vitro using cell free translation systems. In vitro translation systems  
40 derived from both prokaryotic (e.g. *E. coli*) and eukaryotic (e.g. wheat germ, rabbit reticulocytes) cells  
are available and may be chosen based on the expression levels and functional properties of the  
protein of interest. For example, as appreciated by those skilled in the art, in vitro translation is  
required for some display technologies, for example ribosome display. In addition, the library proteins  
may be produced by chemical synthesis methods.

5

*Expression Vectors*

The nucleic acids that encode the antibody library members may be incorporated into an expression vector in order to express the protein. A variety of expression vectors may be utilized to express the library proteins. Expression vectors may comprise self-replicating extra-chromosomal vectors or  
10 vectors which integrate into a host genome. Expression vectors are constructed to be compatible with the host cell type. Thus expression vectors which find use in the present invention include but are not limited to those which enable protein expression in mammalian cells, bacteria, insect cells, and yeast. As is known in the art, a variety of expression vectors are available, commercially or otherwise, that may find use in the present invention for expressing antibody library proteins.

15

Expression vectors typically comprise a library member operably linked with control or regulatory sequences, selectable markers, any fusion partners, and/or additional elements. By "operably linked" herein is meant that the nucleic acid is placed into a functional relationship with another nucleic acid sequence. Generally, these expression vectors include transcriptional and translational regulatory  
20 nucleic acid operably linked to the nucleic acid encoding the library antibody, and are typically appropriate to the host cell used to express the library protein. In general, the transcriptional and translational regulatory sequences may include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. As is also known in the art, expression vectors typically contain a  
25 selection gene or marker to allow the selection of transformed host cells containing the expression vector. Selection genes are well known in the art and will vary with the host cell used.

*Fusion Partners*

Antibody library members may be operably linked to a fusion partner to enable targeting of the  
30 expressed protein, purification, screening, display, and the like. Fusion partners may be linked to the library member sequence via a linker sequences. The linker sequence will generally comprise a small number of amino acids, typically less than ten, although longer linkers may also be used. Typically, linker sequences are selected to be flexible and resistant to degradation. As will be appreciated by those skilled in the art, any of a wide variety of sequences may be used as linkers. For example, a  
35 common linker sequence comprises the amino acid sequence GGGGS.

A fusion partner may be a targeting or signal sequence that directs library antibody protein and any associated fusion partners to a desired cellular location or to the extracellular media. As is known in the art, certain signaling sequences may target a protein to be either secreted into the growth media,  
40 or into the periplasmic space, located between the inner and outer membrane of the cell.

A fusion partner may also be a sequence that encodes a peptide or protein that enables purification and/or screening. Such fusion partners include but are not limited to polyhistidine tags (for example His<sub>6</sub> and His<sub>10</sub> or other tags for use with Immobilized Metal Affinity Chromatography (IMAC) systems

5 (e.g.  $\text{Ni}^{+2}$  affinity columns)), GST fusions, MBP fusions, Strep-tag, the BSP biotinylation target sequence of the bacterial enzyme BirA, and epitope tags which are targeted by antibodies (for example to c-myc tags, flag tags, and the like). As will be appreciated by those skilled in the art, such tags may be useful for purification, for screening, or both. For example, an antibody fragment may be purified using a His-tag by immobilizing it to a  $\text{Ni}^{+2}$  affinity column, and then after purification the same  
 10 His-tag may be used to immobilize the antibody to a  $\text{Ni}^{+2}$  coated plate to perform an ELISA or other binding assay (see "Screening of Library Members" section below).

A fusion partner may enable the use of a selection method to screen antibody library members (see "Screening based on selection methods" below). Fusion partners which enable a variety of selection  
 15 methods are well-known in the art, and all of these find use in the present invention. For example, by fusing the members of an antibody library to the gene III protein, phage display can be used (Kay *et al.*, 1996, *Phage display of peptides and proteins: a laboratory manual*, Academic Press, San Diego, CA); Lowman *et al.*, 1991, *Biochemistry* 30:10832-10838; Smith, 1985, *Science* 228:1315-1317).

Fusion partners may enable antibody library members to be labeled. Alternatively, a fusion partner  
 20 may bind to a specific sequence on the expression vector, enabling the fusion partner and associated antibody library member to be linked covalently or noncovalently with the nucleic acid that encodes them. For example, USSNs 09/642,574; 10/080,376; 09/792,630; 10/023,208; 09/792,626; 10/082,671; 09/953,351; 10/097,100; and 60/366,658; PCTs 00/22906; 01/49058; 02/04852; 02/04853; 02/08023; 01/28702; and 02/07466; all herein expressly incorporated by reference,  
 25 describe such a fusion partner and technique that may find use in the present invention.

#### *Transformation and Transfection Methods*

The methods of introducing exogenous nucleic acid into host cells is well known in the art, and will vary with the host cell used. Techniques include but are not limited to dextran-mediated transfection,  
 30 calcium phosphate precipitation, calcium chloride treatment, polybrene mediated transfection, protoplast fusion, electroporation, viral or phage infection, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei. In the case of mammalian cells, transfection may be either transient or stable.

#### 35 Protein Purification

In a preferred embodiment, antibody library members are purified or isolated after expression. Antibodies may be isolated or purified in a variety of ways known to those skilled in the art. Standard purification methods include chromatographic techniques, including ion exchange, hydrophobic interaction, affinity, sizing or gel filtration, and reversed-phase, carried out at atmospheric pressure or  
 40 at high pressure using systems such as FPLC and HPLC. Purification methods also include electrophoretic, immunological, precipitation, dialysis, and chromatofocusing techniques. Ultrafiltration and diafiltration techniques, in conjunction with protein concentration, are also useful. As is well known in the art, a variety of natural proteins bind antibodies, and these proteins can find use in the present invention for purification of antibody library members. For example, the bacterial

5 proteins A and G bind to the Fc region, and the bacterial protein L binds to the Fab region. Purification can often be enabled by a particular fusion partner. For example, antibody library members may be purified using glutathione resin if a GST fusion is employed, Ni<sup>+2</sup> affinity chromatography if a His tag is employed, or immobilized anti-flag antibody if a flag tag is used. For general guidance in suitable purification techniques, see Protein Purification: Principles and Practice,  
10 3<sup>rd</sup> Ed., Scopes, Springer-Verlag, NY, 1994, hereby expressly incorporated by reference.

The degree of purification necessary will vary depending on the screen or use of the antibody library members. In some instances no purification is necessary. For example in one embodiment, if library antibodies are secreted, screening may take place directly from the media. As is well known in the  
15 art, some methods of selection do not involve purification of library proteins. Thus, for example, if the optimized antibody sequences are made into a phage display library, antibody purification may not be performed.

#### Screening of Library Members

20 Library members may be screened using a variety of methods, including but not limited to those that use in vitro assays, in vivo and cell-based assays, and selection technologies. Automation and high-throughput screening technologies may be utilized in the screening procedures. Screening may employ the use of a fusion partner or label. The use of fusion partners has been discussed above. By "labeled" herein is meant that the antibodies of the invention have one or more elements, isotopes,  
25 or chemical compounds attached to enable the detection in a screen. In general, labels fall into three classes: a) immune labels, which may be an epitope incorporated as a fusion partner that is recognized by an antibody, b) isotopic labels, which may be radioactive or heavy isotopes, and c) small molecule labels, which may include fluorescent and colorimetric dyes, or molecules such as biotin which enable other labeling methods. Labels may be incorporated into the compound at any  
30 position and may be incorporated in vitro or in vivo during antibody expression.

#### *In vitro Assays*

In a preferred embodiment, the functional and/or biophysical properties of antibody library members are screened in an in vitro assay. In vitro assays may allow a broad dynamic range for screening  
35 antibody properties of interest. Properties of library members that may be screened include but are not limited to stability, solubility, and affinity for antigen, antibody receptors, or other proteins which are known to bind the antibody being optimized. Multiple properties may be screened simultaneously or individually. Proteins may be purified or unpurified, depending on the requirements of the assay.

40 In one embodiment, the screen is a qualitative or quantitative binding assay for binding of antibody library members to a protein or nonprotein molecule that is known to bind the antibody. In a preferred embodiment, the screen is a binding assay for measuring the binding of antibody library members to the antibody's antigen. In an alternately preferred embodiment, the screen is an assay for antibody binding to an antibody receptor or some other protein that is known to bind antibodies. For example,

5 a number of proteins are known to bind the Fc region (Ravetch & Bolland, 2001, *Ann. Rev. Immunol.* 19:275-90; Raghavan & Bjorkman, 1996, *Annu. Rev. Cell Dev. Biol.* 12:181-220), including the family of FcγRs, the neonatal receptor FcRn, the complement protein C1q, and the bacterial proteins A and G. Binding assays can be carried out using a variety of methods known in the art. These methods include but are not limited to FRET (Fluorescence Resonance Energy Transfer) and BRET  
10 (Bioluminescence Resonance Energy Transfer) -based assays, AlphaScreen (Amplified Luminescent Proximity Homogeneous Assay), Scintillation Proximity Assay, ELISA (Enzyme-Linked Immunosorbent Assay), SPR (Surface Plasmon Resonance) or BIACORE, isothermal titration calorimetry, differential scanning calorimetry, gel electrophoresis, and chromatography including gel filtration. These and other methods may take advantage of some fusion partner or label of the  
15 antibody library member. Assays may employ a variety of detection methods including but not limited to chromogenic, fluorescent, luminescent, or isotopic labels.

The biophysical properties of antibodies, for example stability and solubility, may be screened using a variety of methods known in the art. Protein stability may be determined by measuring the  
20 thermodynamic equilibrium between folded and unfolded states. For example, antibody library members of the present invention may be unfolded using chemical denaturant, heat, or pH, and this transition may be monitored using methods including but not limited to circular dichroism spectroscopy, fluorescence spectroscopy, absorbance spectroscopy, NMR spectroscopy, calorimetry, and proteolysis. As will be appreciated by those skilled in the art, the kinetic parameters of the folding  
25 and unfolding transitions may also be monitored using these and other techniques. The solubility and overall structural integrity of an antibody may be quantitatively or qualitatively determined using a wide range of methods that are known in the art. Methods which may find use in the present invention for characterizing the biophysical properties of antibody library members include gel electrophoresis, chromatography such as size exclusion chromatography and reversed-phase high performance liquid  
30 chromatography, mass spectrometry, ultraviolet absorbance spectroscopy, fluorescence spectroscopy, circular dichroism spectroscopy, isothermal titration calorimetry, differential scanning calorimetry, analytical ultra-centrifugation, dynamic light scattering, proteolysis, and cross-linking, turbidity measurement, filter retardation assays, immunological assays, fluorescent dye binding assays, protein-staining assays, microscopy, and detection of aggregates via ELISA. Structural  
35 analysis employing X-ray crystallographic techniques and NMR spectroscopy may also find use. In one embodiment, antibody stability and/or solubility may be measured by determining the amount of antibody in solution after some defined period of time. In this assay, the antibody may or may not be exposed to some extreme condition, for example elevated temperature, low pH, or the presence of denaturant. Because antibody function typically requires a stable, soluble, and/or well-  
40 folded/structured antibody, the functional (i.e. binding) assays described above also provide a way to perform such an assay. For example, a solution comprising an antibody variant could be assayed for its ability to bind antigen, then exposed to elevated temperature for one or more defined periods of time, then assayed for antigen binding again. Because unfolded and aggregated antibody is not

- 5 expected to be capable of binding antigen, the amount of antibody activity remaining provides a measure of the antibody variant's stability and solubility.

*In Vivo or Cell-based Assays*

10 In a preferred embodiment, the library is screened using one or more cell-based or in vivo-based assays. Cell types for such assays may be prokaryotic or eukaryotic. For such assays, antibody library members, purified or unpurified, are typically added exogenously such that cells are exposed to individual variants or pools of variants belonging to a library. These assays are typically, but not always, based on the function of the antibody, that is the ability of the antibody to bind an antigen and/or some protein which naturally binds the antibody, for example an Fc receptor. Such assays  
15 often involve monitoring the response of cells to antibody, for example cell survival, cell death, change in cellular morphology, or transcriptional activation such as cellular expression of a natural gene or reporter gene. For example, anti-cancer antibodies may cause apoptosis of certain cell lines expressing the antibody's target antigen, or they may mediate attack on target cells by immune cells which have been added to the assay. Methods for monitoring cell death or viability are known in the art, and include the use of dyes, immunochemical, cytochemical, or radioactive reagents. For  
20 example, caspase staining assays may enable apoptosis to be measured, and uptake of radioactive substrates or the dye alamar blue may enable cell growth or activation to be monitored. Transcriptional activation may also serve as a method for assaying antibody function in cell-based assays. In this case, response may be monitored by assaying for natural genes or proteins which  
25 may be upregulated, for example the release of certain interleukins may be measured, or alternatively readout may be via a reporter construct. Cell-based assays may also involve the measure of morphological changes of cells as a response to the presence of an antibody library variant.

Alternatively, cell-based screens are performed directly using cells that have been transformed or  
30 transfected with nucleic acids encoding antibody library members. That is, antibody library variants are not added exogenously to the cells. For example, in one embodiment, the cell-based screen utilizes cell surface display. A fusion partner can be employed that enables display of antibodies on the surface of cells (Wittrup, 2001, *Curr. Opin. Biotechnol.*, 12:395-399). Cell surface display methods which may find use in the present invention include but are not limited to display on bacteria  
35 (Georgiou *et al.*, 1997, *Nat. Biotechnol.* 15:29-34.; Georgiou *et al.*, 1993, *Trends Biotechnol.* 11:6-10; Lee *et al.*, 2000, *Nat. Biotechnol.* 18:645-648; Jun *et al.*, 1998, *Nat. Biotechnol.* 16:576-80.), yeast (Boder & Wittrup, 2000, *Methods Enzymol.* 328:430-44; Boder & Wittrup, 1997, *Nat. Biotechnol.* 15:553-557), and mammalian cells (Whitehorn *et al.*, 1995, *Bio/technology* 13:1215-1219). In an alternate embodiment, antibodies are not displayed on the surface of cells, but rather are screened  
40 intracellularly or in some other cellular compartment. For example, periplasmic expression and cytometric screening (Chen *et al.*, 2001, *Nat. Biotechnol.*, 19: 537-542), the protein fragment complementation assay (Johnsson & Varshavsky, 1994, *Proc. Natl. Acad. Sci. USA*, 91: 10340-10344.; Pelletier *et al.*, 1998, *Proc. Natl. Acad. Sci. USA* 95:12141-12146), and the yeast two hybrid screen (Fields & Song, 1989, *Nature* 340:245-246) may find use in the present invention.

5 Alternatively, if the antibody imparts some selectable growth advantage to a cell, this property may be used to screen or select for antibody variants.

The biological properties of one or more antibody library members, including clinical efficacy, pharmacokinetics, and toxicity, may also be characterized in cell, tissue, and whole organism  
10 experiments.

#### *Screening Based on Selection Methods*

As is known in the art, a subset of screening methods are those that select for favorable members of a library. Said methods are herein referred to as "selection methods", and these methods find use in  
15 the present invention for screening antibody libraries. When antibody libraries are screened using a selection method, only those members of a library which are favorable, that is which meet some selection criteria, are propagated, isolated, and/or observed. As will be appreciated, because only the most fit antibody variants are observed, such methods enable the screening of libraries which are larger than those screenable by methods which assay the fitness of library members individually.

20 Selection is enabled by any method, technique, or fusion partner which links, covalently or noncovalently, the phenotype of an antibody variant with its genotype, that is the function of an antibody with the nucleic acid that encodes it. For example the use of phage display as a selection method is enabled by the fusion of library members to the gene III protein. In this way, selection or isolation of antibody proteins which meet some criteria, for example binding affinity for antigen, also  
25 selects for or isolates the nucleic acid which encodes it. Once isolated, the gene or genes encoding library antibody variants may then be amplified. This process of isolation and amplification, referred to as panning, may be repeated, allowing favorable antibody variants in the library to be enriched. Nucleic acid sequencing of the attached nucleic acid ultimately allows for gene identification.

30 A variety of selection methods are known in the art which may find use in the present invention for screening antibody libraries. These include but are not limited to phage display (Phage display of peptides and proteins: a laboratory manual, Kay *et al.*, 1996, Academic Press, San Diego, CA; Lowman *et al.*, 1991, *Biochemistry* 30:10832-10838; Smith, 1985, *Science* 228:1315-1317) and its derivatives such as selective phage infection (Malmberg *et al.*, 1997, *J. Mol. Biol.* 273:544-551),  
35 selectively infective phage (Krebber *et al.*, 1997, *J. Mol. Biol.* 268:619-630), and delayed infectivity panning (Benhar *et al.*, 2000, *J. Mol. Biol.* 301:893-904), cell surface display (Wittrup, 2001, *Curr. Opin. Biotechnol.*, 12:395-399) such as display on bacteria (Georgiou *et al.*, 1997, *Nat. Biotechnol.* 15:29-34.; Georgiou *et al.*, 1993, *Trends Biotechnol.* 11:6-10; Lee *et al.*, 2000, *Nat. Biotechnol.* 18:645-648; Jun *et al.*, 1998, *Nat. Biotechnol.* 16:576-80), yeast (Boder & Wittrup, 2000, *Methods Enzymol.* 328:430-44; Boder & Wittrup, 1997, *Nat. Biotechnol.* 15:553-557), and mammalian cells  
40 (Whitehorn *et al.*, 1995, *Bio/technology* 13:1215-1219), as well as in vitro display technologies (Amstutz *et al.*, 2001, *Curr. Opin. Biotechnol.* 12:400-405) such as polysome display (Mattheakis *et al.*, 1994, *Proc. Natl. Acad. Sci. USA* 91:9022-9026), ribosome display (Hanes *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:4937-4942), mRNA display (Roberts & Szostak, 1997, *Proc. Natl. Acad. Sci. USA*

- 5 94:12297-12302; Nemoto *et al.*, 1997, *FEBS Lett.* 414:405-408), and ribosome-inactivation display system (Zhou *et al.*, 2002, *J. Am. Chem. Soc.* 124, 538-543)

Other selection methods which may find use in the present invention include methods that do not rely on display, such as *in vivo* methods including but not limited to periplasmic expression and cytometric  
 10 screening (Chen *et al.*, 2001, *Nat. Biotechnol.*, 19: 537-542), the protein fragment complementation assay (Johnsson & Varshavsky, 1994, *Proc. Natl. Acad. Sci. USA*, 91: 10340-10344; Pelletier *et al.*, 1998, *Proc. Natl. Acad. Sci. USA* 95:12141-12146), and the yeast two hybrid screen (Fields & Song, 1989, *Nature* 340:245-246) used in selection mode (Visintin *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96: 11723-11728). In an alternate embodiment, selection is enabled by a fusion partner which binds  
 15 to a specific sequence on the expression vector, thus linking covalently or noncovalently the fusion partner and associated antibody library member with the nucleic acid that encodes them. In an alternative embodiment, *in vivo* selection can occur if expression of the library antibody imparts some growth, reproduction, or survival advantage to the cell.

20 As is known in the art, a subset of selection methods referred to as "directed evolution methods" are those that include the mating or breeding of favorable sequences during selection, sometimes with the incorporation of new mutations. As will be appreciated by those skilled in the art, directed evolution methods can facilitate identification of the most favorable sequences in a library, and can increase the diversity of sequences that are screened. A variety of directed evolution methods are known in the art  
 25 that may find use in the present invention for screening antibody libraries, including but not limited to DNA shuffling (WO 00/42561 A3; WO 01/70947 A3), exon shuffling (US 6365 377 B1; Kolkman & Stemmer, 2001, *Nat. Biotechnol.* 19:423-428), family shuffling (Crameri *et al.*, 1998, *Nature* 391:288-291; US 6376246 B1), RACHITT™ (Coco *et al.*, 2001, *Nat. Biotechnol.* 19:354- 359; WO 02/06469 A2), STEP and random priming of *in vitro* recombination (Zhao *et al.*, 1998, *Nat. Biotechnol.* 16:258-  
 30 261; Shao *et al.*, 1998, *Nucleic Acids Res.* 26:681-683), exonuclease mediated gene assembly (US 6352842 B1; US 6361974 B1), Gene Site Saturation Mutagenesis™ (US 6358709 B1), Gene Reassembly™ (US 6358709B1), SCRATCHY (Lutz *et al.*, 2001, *Proc. Natl. Acad. Sci. USA* 98:11248-11253), DNA fragmentation methods (Kikuchi *et al.*, *Gene* 236:159-167), and single-stranded DNA shuffling (Kikuchi *et al.*, 2000, *Gene* 243:133-137), all of which are herein expressly incorporated by  
 35 reference.

### Design Strategies

A variety of computational screening design strategies are provided for optimization of the physico-chemical properties of antibodies, including stability, solubility, and antigen binding affinity. These  
 40 strategies can be used individually or in combination.

### Stability Optimization

There is frequently a need to enhance the stability of an antibody. Lower stability of a full-length antibody or an antibody fragment may result in greater amount of nonnative and thus nonfunctional

5 species, increased susceptibility to degradation, and greater tendency for aggregation. Increased degradation and aggregation may result in lower in vivo half-life of the molecule if the antibody is a therapeutic, further decreasing activity.

10 In one object of the present invention, computational screening methodology is used to enhance the stability of an antibody. A number of design strategies are disclosed for antibody stabilization, including strategies which employ experimental information and/or sequence information to guide choice of variable positions, choice of amino acids considered at those positions, and/or generation of one or more experimental libraries from computational output. The disclosed design strategies are not meant to constrain the present invention to any particular application or theory of operation.

15 Rather, the present invention relates as novel not only these provided individual strategies, but the general use of computational screening to enhance the stability of antibodies.

The stability of an antibody is comprised of: a) the stabilities of each individual Ig domain which make up the antibody, and b) the stabilities or affinities of interdomain interactions if the antibody is

20 composed of more than one Ig domain. Thus two main strategies for utilizing computational screening methodology to stabilize antibodies are to enhance the stability of individual Ig domains, and enhance interface stability between individual Ig domains.

#### *Domain Stability*

25 The stability of an antibody is determined in part by the individual stabilities of each of the Ig domains that comprise it. In one embodiment, computational screening is used to stabilize an antibody by enhancing the stability of one or more individual Ig domains. In this embodiment, more favorable interactions are designed within one or more individual Ig domains, thereby increasing the global stability of the antibody as a whole. For an antibody which is made up of more than one Ig domain,

30 each individual Ig domain may be engineered for greater stability. Thus for example, for antibodies derived from human, mouse, rat, or rabbit antibodies, the stability may be improved by stabilizing one or more of domains V<sub>H</sub>, V<sub>L</sub>, C<sub>γ</sub>1, C<sub>L</sub>, C<sub>γ</sub>2, and C<sub>γ</sub>3.

In one embodiment, the interior of an Ig domain or Ig domains are redesigned to be more stable. For

35 example, as will be appreciated by those skilled in the art, the van der Waals packing interactions between nonpolar residues in the core play an important role protein stability. Mutations may be designed that result in more favorable interactions between interior residues. In another embodiment, non-interior residues, that is boundary or surface positions an Ig domain or domains are designed to be more stable. For example, greater stability may be gained when amino acid side chains which

40 have the capacity to donate a hydrogen bond are interacting with a molecule which is capable of accepting a hydrogen bond, whether this molecule be another side chain, the protein backbone, or solvent. Interior and non-interior residues may be identified by objective methods such as degree of solvent exposure, as described above, subjective methods such as visual inspection by one skilled in the art of protein structural biology, or other methods. As described above, variable positions and

- 5 amino acids considered at those positions may be chosen using any variety of approaches, including but not limited to approaches based on solvent exposure, approaches which are hypothesis-driven, approaches which utilize experimental information, approaches which utilize sequence information, or any combination of these and other approaches.
- 10 A number of examples are provided below which describe the use of computational screening methods to stabilize the Ig domains of an antibody. These examples are not meant to constrain the present invention to any particular application or theory of operation. Rather, the present invention relates as novel not only these provided individual examples, but the general use of computational screening methodology to enhance the stability of an Ig domain or Ig domains in order to optimize an
- 15 antibody for greater stability.

#### *Interface Stability*

- The stability of multi-Ig domain antibodies, that is to say full-length antibodies and antibody fragments which are composed of more than one Ig domain, are determined in part by the affinities of the
- 20 interactions between domains (Worn & Plückthun, 2001, *J. Mol. Biol.* 305:989-1010). Two interacting Ig domains exist in equilibrium between bound and unbound states. In the unbound state, Ig domains have a greater tendency to unfold and aggregate than when they are in the bound state. Thus by designing more favorable interactions between residues that mediate the interdomain interaction, the bound state may be stabilized, thereby stabilizing the antibody as a whole. In one embodiment of the
- 25 present invention, computational screening is used to engineer mutations that result in more favorable interactions between individual Ig domains. As shown in Figure 1, for human antibodies there are five interdomain interfaces that may be optimized using computational screening methodology:  $V_H-V_L$ ,  $C\gamma 1-C_L$ ,  $V_H-C\gamma 1$ ,  $V_L-C_L$ , and  $C\gamma 3-C\gamma 3$ . The stability of a Fab is dependent on the interactions at only a subset of these interfaces:  $V_H-V_L$ ,  $C\gamma 1-C_L$ ,  $V_H-C\gamma 1$ , and  $V_L-C_L$ .

- 30 Greater interdomain stability may be obtained by engineering more energetically favorable interactions between residues that mediate the interdomain interface. Such designed interactions could involve more favorable packing interactions, hydrogen bond interactions, electrostatic interactions, hydrophobic interactions, and the like. Interface residues may be identified by objective
- 35 methods such as degree of solvent exposure, as described above, subjective methods such as visual inspection by one skilled in the art of protein structural biology, or other methods. As described above, variable positions and amino acids considered at those positions may be chosen using any variety of approaches, including but not limited to approaches based on solvent exposure, approaches which are hypothesis-driven, approaches which utilize experimental information,
- 40 approaches which utilize sequence information, or any combination of these and other approaches.

In one embodiment, the interface is designed to have more favorable nonpolar interactions, for example by engineering the interface with more nonpolar volume than that in the antibody template, by designing nonpolar residues which pack better together than that in the antibody template, and the

5 like. As will be appreciated by those skilled in the art, this may be thought of as the interface version of a redesigned hydrophobic core. Here, however, variable positions are those that make up the interface between Ig domains instead of the core of an Ig domain. In an alternate embodiment, the interface is designed to have more favorable polar interactions, for example by engineering the interface with more polar amino acids than that in the antibody template, by designing nonpolar  
10 residues with more optimized hydrogen bonds, electrostatic interactions, and the like. As will be appreciated by those in the art, greater polar character at the interface may enable the bound/unbound equilibrium between Ig domains to be more reversible. In the unbound state, the residues which make up the interface with the other Ig domain and are normally sequestered from solvent become exposed to solvent. Nonpolar residues have a higher tendency to aggregate than  
15 polar residues, and therefore greater nonpolar character at the interdomain interface may result in a greater tendency to aggregate in the unbound form, resulting in non-reversibility of the unbinding/binding transition. Irreversible aggregation means that the antibody cannot get back to its native bound state (i.e. the Ig domain interface is not reformed). This property of Ig domain interfaces in antibodies is supported experimentally (Worn & Plückthun, 2001, *J. Mol. Biol.* 305:989-1010; Ewert  
20 *et al.*, 2002, *Biochemistry*, 41:3628-3636). In an alternate embodiment, the interface is engineered with more favorable nonpolar and polar interactions.

A number of examples are provided below in which describe the use of computational screening methods to stabilize the interfaces between Ig domains. These examples illustrate how a variety of  
25 interactions may be designed at interdomain interfaces that result in greater stability. These examples are not meant to constrain the present invention to any particular application or theory of operation. Rather, the present invention relates as novel not only these provided individual examples, but the general use of computational screening methodology to design more energetically favorable inter-Ig domain interactions in order to stabilize an antibody.

30

#### Solubility Optimization

There is frequently a need to enhance the solubility of an antibody. Lower solubility of an antibody may result in a greater fraction of nonfunctional species, increased susceptibility to degradation, and shorter in vivo half-life and lower efficacy if the antibody is a therapeutic. Poor solubility may also  
35 place severe constraints on antibody formulation and route of administration. A number of design strategies are suggested for using computational screening methods to enhance the solubility of an antibody, all of which are embodiments of the present invention.

In one embodiment, surface exposed nonpolar residues in an antibody are replaced with polar  
40 residues which are predicted by computational screening calculations to be favorable. Underlying this strategy is the principle that polar residues are more soluble than nonpolar ones. This principle is well known in the art. In regard to which residues are more polar or nonpolar than others, such a judgment may be made subjectively or objectively. Subjectively, for example, one skilled in the art of protein structural biology appreciates qualitatively that amino acids such as leucine, tryptophan, and

5 methionine are more nonpolar, and thus potentially more prone to cause aggregation when exposed to solvent, than amino acids such as serine, asparagine, and glutamate. Objective and quantitative measurements of hydrophobicity are also known in the art. For example, the free energies of transfer of an amino acid from non-aqueous to aqueous solution have been used to generate relative rankings of amino acid hydrophobicity, and such methods find use in the present invention. Variable positions  
10 and amino acids considered at those positions may be chosen using any variety of approaches, as described above, including but not limited to approaches based on solvent exposure, approaches which are hypothesis-driven, approaches which utilize experimental information, approaches which utilize sequence information, or any combination of these and other approaches.

15 A number of strategies for replacing exposed nonpolar amino acids find use in the present invention. In one embodiment, residues which may be replaced include residues which are exposed to solvent on individual Ig domains, or which lie at the interface between Ig domains. In this regard, all Ig domains of a human antibody, including V<sub>H</sub>, V<sub>L</sub>, C<sub>γ</sub>1, C<sub>L</sub>, C<sub>γ</sub>2, and C<sub>γ</sub>3, as well as the linkers and/or hinges which connect them, have surface residues which could be replaced with amino acids which  
20 may impart greater solubility to the antibody. In another embodiment, variable positions reside in a region of an antibody fragment which in the context of a full-length antibody or larger antibody fragment makes up the interface with another Ig domain. As will be appreciated by those skilled in the art, antibody fragments are generated by removing certain regions or domains of an antibody. As a result, regions of an Ig domain which interact with another Ig domain in the larger antibody may  
25 become exposed to solvent in the context of an antibody fragment. For example, the V<sub>H</sub> and V<sub>L</sub> residues which make up the V<sub>H</sub>/C<sub>γ</sub>1 and V<sub>L</sub>/C<sub>L</sub> interfaces of an antibody are exposed to solvent in an scFv fragment of that antibody (Nieba *et al.*, 1997, *Protein Eng.* 10:435-44). The result for an scFv, or any other antibody fragment, may be increased propensity for aggregation and thus lower solubility. Computational screening methods may be used to engineer mutations at these positions which result  
30 in greater solubility of the antibody fragment.

Several additional strategies may also be used to optimize solubility. For example, it is known in the art that protein solubility is typically lowest when the pH of the solution is equal to the isoelectric point (pI) of the protein. Under such conditions, the net charge of the protein is equal to zero. It is possible  
35 to optimize solubility by altering the number and location of ionizable residues in the antibody to adjust the pI. In other cases, improvements in solubility may result from optimizing the stability of the antibody, as discussed above. As is well known in the art, proteins are much more prone to aggregation in unfolded or partially folded states. Thus proteins that are well folded, structured, and/or stable are typically more soluble. Accordingly, computational screening which stabilizes an  
40 antibody, for example by one or more design strategies discussed above, may also be used to enhance antibody solubility. Additionally, if the antibody contains one or more cysteines that do not form disulfide bonds in the native antibody structure, replacing such cysteines with less reactive, structurally compatible residues can prevent the formation of unwanted intra- and inter-molecular

5     disulfide bonds. As will be appreciated by those skilled in the art, additional strategies could also be used to optimize the solubility of antibodies.

#### Affinity Maturation

10     There is frequently a need to enhance the affinity of an antibody for its antigen. This process is referred to as affinity maturation, and following this process, the antibody may then be said to be affinity matured. The binding affinity of an antibody for its target is a critical parameter for its success as a therapeutic, diagnostic, or reagent. Higher affinity for antigen may result in a more efficacious antibody therapeutic. As discussed above, enhancement of antigen affinity is frequently wanted or needed for a variety of forms and sources of antibodies such as those that are substantially human,  
15     nonhuman, chimeric, or humanized. A particular case which demands affinity maturation is subsequent to humanization. As discussed above, this technique to reduce the immunogenicity of antibody therapeutics often results in loss of binding affinity for antigen, and thus regaining this affinity is typically desired.

20     Computational screening methods may be applied to antibody affinity maturation using a number of design strategies, all of which are embodiments of the present invention. Strategies for affinity maturation include but are not limited to those which use only a structure or structures of bound antibody/antigen complexes, only a structure or structures of unbound antibodies, or structures of both bound and unbound antibody. These strategies need not be defined by the structural information  
25     that is available, but rather may be defined by the structural information that is employed. For example, to affinity mature an antibody it may be useful to carry out design calculations on an unbound antibody template that is a structure of the antibody alone without antigen, even though a structure of the antibody/antigen complex may be available. The structure of the unbound antibody may be available, or could be obtained by deleting antigen coordinates from the structure of the  
30     complex.

As discussed above, antibody templates may be obtained from a variety of sources, including but not limited to X-ray crystallographic techniques, NMR techniques, de novo modeling, and homology modeling. Antibody/antigen complexes may furthermore be obtained using docking methods. For  
35     example, if the antibody/antigen complex structure is not available, it may be modeled by docking the antigen into the antibody variable region. Methods for this process are known in the art. Variable positions and amino acids considered at those positions may be chosen using any variety of approaches, as described above, including but not limited to approaches based on solvent exposure, approaches which are hypothesis-driven, approaches which utilize experimental information,  
40     approaches which utilize sequence information, or any combination of these and/or other approaches.

In one embodiment, computational screening is used to affinity mature an antibody by using the structure of a bound antibody/antigen complex as the template for design calculations. In this strategy, one or more antibody mutations are design that result in more favorable interactions (i.e.,

5 higher affinity) between the antibody and its antigen. In one embodiment, only antibody residues which directly contact antigen, referred to herein as "contact residues" are allowed to vary in design calculations. In an alternate embodiment, variable antibody positions may include residues which do not contact antigen, alone or in addition to residues which do contact antigen. For example, the variable positions in a design calculation could be set to those residues which interact with contact  
10 residues, but are not themselves contact residues. As will be appreciated by those skilled in the art, the subtle conformations of contact residues which are optimal for antigen binding are determined in part by the conformations of the surrounding residues. By using computational screening to explore substitutions in the shell of residues which interact with contact residues, a quality diversity of new contact residue conformations may be sampled. In an alternate embodiment, contact residues and  
15 residues which are not contact residues are variable positions in design calculations.

In another embodiment, computational screening is used to affinity mature an antibody by using the structure of an uncomplexed antibody structure, i.e. a structure of an antibody which is not bound to its antigen, as the template for design calculations. In this strategy, antibody residues which contact  
20 antigen or which are believed to contact antigen are mutated to residues which are energetically favorable in the context of the structural template. The primary goal of this approach is to generate quality diversity within an experimental library such that the distribution within the library is skewed towards a larger percentage of variants which are energetically compatible with the antibody than would be expected if variants were designed randomly. Although the antibody variants in this library  
25 are not directly computationally screened to possess higher affinity for antigen, such variants will likely still be present in the library. The use of computational screening enables the vast sequence space of mutations which are inconsistent with the antibody structure to be trimmed from the library, thereby increasing the chances of finding in an experimental screen those variants which possess higher antigen binding affinity. In the absence of an antibody/antigen complex structure, it is not possible to  
30 identify contact residues by visual inspection. Thus, experimental and sequence information are particularly useful in this case, as these may provide insight into which residues are important determinants of antigen binding.

In another embodiment, computational screening methods are used to affinity mature an antibody by  
35 combining results from design calculations which use the structures of both a bound antibody/antigen complex and an unbound antibody structure as templates for design calculations. In one embodiment, computational screening is used to engineer mutations at or near the antibody/antigen interface that are energetically favorable in the context of both the bound and unbound antibody structures. For this strategy, output from two sets of design calculations could be used to generate an  
40 experimental library. For example, one set of calculations could involve those which use one or more unbound antibody structures as the template(s), and another set of calculations could use one or more bound antibody/antigen structures as the template(s). The experimental library could be comprised of variants which are predicted to be energetically favorable in both sets of calculations. In one embodiment, variants which are predicted to be energetically favorable in both structures are

5 included in the library. In an alternate embodiment, variants which are predicted to be energetically favorable in at least one of the structures are included in the library. As is illustrated in the examples below, it is a preferred embodiment to have at least one of the variable regions located in a framework region, a complementarity determining region or a combination of both regions.

10 A number of examples are provided below which describe the use of computational screening to affinity mature antibodies. These examples are not meant to constrain the present invention to any particular application or theory of operation. Rather, the present invention relates as novel not only these provided individual examples, but the general use of computational screening methods to affinity mature antibodies.

15

### EXAMPLES

A number of examples are provided below to illustrate implementation of the design strategies discussed above to optimize antibodies. These examples employ a variety of strategies, approaches, methods, and so forth to choose variable positions, choose amino considered at those positions, calculate energies, search sequence space using optimization algorithms, and generate experimental libraries. Libraries generated from these examples could be subsequently screened experimentally to obtain optimized antibody variants, become part of other libraries which could be subsequently screened experimentally, or serve other purposes. These examples are not meant to constrain the present invention to any particular application or theory of operation. Rather, the present invention relates as novel not only to these provided individual examples, but the general use of computational screening to enhance antibody stability, improve antibody solubility, and increase the affinity of antibodies for antigen.

30 Figure 3 shows a list of the antibody structures which are used as templates in the provided examples. Unless otherwise noted, the groups of core, surface, and boundary for choice of amino acids considered at variable positions are composed of the following sets of amino acids: core = alanine, valine, isoleucine, leucine, phenylalanine, tyrosine, tryptophan, and methionine; surface = alanine, serine, threonine, aspartic acid, asparagine, glutamine, glutamic acid, arginine, lysine and histidine; boundary = alanine, serine, threonine, aspartic acid, asparagine, glutamine, glutamic acid, arginine, lysine, histidine, valine, isoleucine, leucine, phenylalanine, tyrosine, tryptophan, and methionine; All or All 20 = all 20 natural amino acids.

#### Stability Optimization

40 As discussed above, two main strategies for utilizing computational screening methodology to stabilize antibodies are to enhance the stability of individual Ig domains, and enhance interface stability between individual Ig domains.

5    Domain Stability

The stability of an antibody can be increased by designing more favorable interactions within one or more individual Ig domains. For an antibody which is made up of more than one Ig domain, each individual Ig domain can be engineered for greater stability. Thus for example, for a human, mouse, rat, or rabbit antibody, stability can be improved by stabilizing one or more of domains V<sub>H</sub>, V<sub>L</sub>, C<sub>γ</sub>1, C<sub>L</sub>,  
10   C<sub>γ</sub>2, and C<sub>γ</sub>3.

*Example 1: Campath V<sub>H</sub> Domain Stabilization*

The heavy chain variable domain (V<sub>H</sub>) of Campath was stabilized using computational screening methods to design more favorable interactions within the interior of the protein. Campath is a  
15   humanized antibody that is currently marketed for treatment for B-cell chronic lymphocytic leukemia. The high resolution structure is available of the complex of the Campath Fab with its target antigen, a peptide from the cell surface protein CD52. This structure, PDB accession code 1CE1, served as the template for design calculations. The V<sub>H</sub> domain of Campath, and most antibodies, has an extensive interior which is critical to its stability. This interior can be thought of as being made up  
20   of two separate hydrophobic cores which are separated by the central disulfide bond. These cores are referred to as the upper core and lower core, with the directional distinction being defined when the CDRs are facing upward as shown in Figure 4. As will be appreciated by those skilled in the art, packing interactions between the hydrophobic residues which make up these cores play a key role in V<sub>H</sub> stability, and thus in the stability of any antibody to which V<sub>H</sub> belongs. Computational screening  
25   was applied to design more stable packing interactions in the V<sub>H</sub> lower core. Variable positions were chosen by visual inspection of the 1CE1 structure, and these positions are shown in Figure 4 and listed in Figure 5a. Because these positions are almost completely sequestered from solvent, the amino acids considered were chosen as the set belonging to the core classification. The conformations of amino acids at variable positions were represented as a set of backbone-  
30   independent side chain rotamers derived from the rotamer library of Dunbrack & Cohen (Dunbrack & Cohen, 1997, *Protein Science* 6:1661-1681).

The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation,  
35   electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This ground state, and the WT Campath sequence, are shown in Figure 5a. The fact that the ground state is very similar to the WT sequence validates the computational screening method. As will be appreciated by those in the art, the predicted lowest energy sequence is not necessarily the true lowest energy sequence because of errors, primarily in  
40   the scoring function, coupled with the fact that subtle conformational differences in proteins can result in dramatic differences in stability. However, the predicted ground state sequence is likely to be close to the true ground state, and thus this problem can be hedged by screening variants close in sequence space and in energy around the predicted ground state. Towards this goal, in order to generate a diversity of sequences for an experimental library, a Monte Carlo algorithm was used to

- 5 evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 5a shows the output sequence lists from this Monte Carlo search.

These results can be used to generate one or more experimental libraries which can be screened for increased antibody stability. As discussed above, there are a variety of ways to generate an  
10 experimental library. Library 1, shown in Figure 5b is a defined library of just the ground state sequence. Library 2, shown in Figure 5c, is a combinatorial library in which a 1% cutoff of occupancy has been applied to the Monte Carlo output, that is to say that only amino acid substitutions which occur in 10 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. Because valine does not occur at heavy chain position 117 in the Monte Carlo output, the WT  
15 sequence is not represented. It may be judicious to include this valine at 117 H so that the WT amino acids are represent combinatorially in library 2. The combination of all of these substitutions with all other substitutions results in a combinatorial complexity of 864, i.e. there are 864 possible variants in the library.

20 *Example 2: Campath V<sub>H</sub> Domain Stabilization*

The light chain variable domain (V<sub>L</sub>) of Campath was also stabilized by using computational screening methods. Like the V<sub>H</sub> domain, V<sub>L</sub> has an extensive interior which can be thought of as being made up of an upper and lower core, separated by the central disulfide bond, shown in Figure 6. Computational screening was applied to design more stable packing interactions in the V<sub>L</sub> upper core.  
25 Stabilization of the upper core may be less straightforward than the lower core because subtle conformational changes to the upper may more directly impact the conformation of the CDRs, and thus mutations may affect antigen binding. Variable positions were chosen by visual inspection of the 1CE1 structure, and these positions are shown in Figure 6 and listed in Figure 7a. For most variable positions, the amino acids conserved were chosen as the set belonging to the core classification  
30 because they are sequestered from solvent. Substitutions at two light chain positions, 92 and 97, could potentially make favorable polar interactions, and so amino acids considered for these positions were chosen as the set belonging to the boundary classification. The conformations of amino acids at variable-positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library.

35 The CE1 structure was used as the template for design calculations. The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This  
40 ground state, and the WT Campath sequence, are shown in Figure 7a. The fact that the WT sequence is predicted to be the ground state validates the computational screening method. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 7a shows the output sequence lists from this Monte Carlo search.

5 These results can be used to generate one or more experimental libraries which can be subsequently screened for increased antibody stability. An experimental library, shown in Figure 7b, was derived from this set of designed calculations by applying a 5% cutoff of occupancy to the Monte Carlo output, i.e. only amino acid substitutions which occur in 50 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. This combinatorial library has a complexity of 448.

10 *Example 3: Campath C $\gamma$ 1 Domain Stabilization*

The heavy chain constant domain 1 (C $\gamma$ 1) is also important to antibody stability. This domain is a part of the antibody constant region, and thus improvements made are widely applicable to antibodies, independent of what antigen is bound at the variable region. The C $\gamma$ 1 of Campath was stabilized  
15 using computational screening methods to design more favorable interactions within the interior of the protein. Like most immunoglobulin domains, C $\gamma$ 1 has an extensive interior made up of an upper and lower core, separated by the central disulfide bond, shown in Figure 8. Computational screening was applied to design more stable packing interaction in the C $\gamma$ 1 upper core. Variable positions were chosen by visual inspection of the 1CE1 structure, and these positions are shown in Figure 8 and  
20 listed in Figure 9a. The majority of the chosen core variable positions are sequestered from solvent, and therefore the amino acids conserved were chosen as the set belonging to the core classification. The exception is heavy chain position 173, substitutions at which could potentially make favorable polar interactions, and so amino acids considered for this position were chosen as the set belonging to the boundary classification. The conformations of amino acids at variable positions were  
25 represented as a set of side chain rotamers derived from a backbone-independent rotamer library. The CE1 structure was used as the template for design calculations. The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This  
30 ground state, and the WT Campath sequence, are shown in Figure 9a. The fact that the predicted ground state sequence is very similar to the WT sequence validates the computational screening method. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 9a shows the output sequence lists from this Monte Carlo search.

35 These results can be used to generate one or more experimental libraries which can be subsequently screened for increased antibody stability. An experimental library, shown in Figure 9b, was derived from this set of designed calculations by applying a 5% cutoff of occupancy to the Monte Carlo output, i.e. only amino acid substitutions which occur in 50 or greater variants out of the 1000 Monte Carlo  
40 output sequences are included in the library. This combinatorial library has a complexity of 192.

*Example 4: Fc C $\gamma$ 2 Domain Stabilization*

The heavy chain constant domain 2 (C $\gamma$ 2) is also important to antibody stability. This domain is part of the antibody Fc region, and thus improvements made are widely applicable to antibodies,

5 independent of what antigen is bound at the variable region. The Fc C $\gamma$ 2 domain was stabilized using computational screening methods to design more favorable interactions within the interior of the protein. The high resolution structure of human Fc has been solved. This structure, PDB accession code 1DN2, served as the template for design calculations. Like most immunoglobulin domains, C $\gamma$ 2 has an extensive interior made up of an upper and lower core, separated by the central disulfide  
10 bond, shown in Figure 10. Computational screening was applied to design more stable packing interactions in the C $\gamma$ 2 upper core. Variable positions were chosen by visual inspection of the 1DN2 structure, and these positions are shown in Figure 10 and listed in Figure 11a. The majority of the chosen core variable positions are sequestered from solvent, and therefore the amino acids conserved were chosen as the set belonging to the core classification. The exception is position 332,  
15 substitutions at which could potentially make favorable polar interactions, and so amino acids considered for this position were chosen as the set belonging to the boundary classification. The conformations of amino acids at variable positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library.

20 The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This ground state, and the WT Fc sequence, are shown in Figure 11a. The fact that the predicted ground state sequence is very similar to the WT sequence validates  
25 the computational screening method. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 11a shows the output sequence lists from this Monte Carlo search.

30 These results can be used to generate one or more experimental libraries which can be screened for increased antibody stability. An experimental library, shown in Figure 11b, was derived directly from this set of designed calculations, i.e. no cutoff criteria were applied. This combinatorial library has a complexity of 336.

#### 35 *Example 5: Fc C $\gamma$ 3 Domain Stabilization*

The heavy chain constant domain 3 (C $\gamma$ 3) is also important to antibody stability. This domain is part of the antibody Fc region, and thus improvements made are widely applicable to antibodies, independent of what antigen is bound at the variable region. The Fc C $\gamma$ 3 domain was stabilized by using computational screening methods to design more favorable interactions within the interior of the  
40 protein. Like most immunoglobulin domains, C $\gamma$ 2 has an extensive interior made up of an upper and lower core, separated by the central disulfide bond, shown in Figure 12. Computational screening was applied to design more stable packing interaction in the C $\gamma$ 3 lower core. Variable positions were chosen by visual inspection of the 1DN2 structure, and these positions are shown in Figure 12 and listed in Figure 13a. The majority of the chosen core variable positions are sequestered from solvent,

5 and therefore the amino acids conserved were chosen as the set belonging to the core classification. The exceptions are positions 358 and 391, substitutions at which could potentially make favorable polar interactions, and so amino acids considered for these positions were chosen as the set belonging to the boundary classification. The conformations of amino acids at variable positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library.

10 1DN2 was used as the structural template for design calculations. The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This ground state, and the WT Fc sequence, are shown in Figure 13a. The fact that the predicted ground

15 state sequence is very similar to the WT sequence validates the computational screening technology. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 13a shows the output sequence lists from this Monte Carlo search.

20 These results can be used to generate one or more experimental libraries which can be screened for increased antibody stability. An experimental library, shown in Figure 13b, was derived from this set of designed calculations by applying a 1% cutoff of occupancy to the Monte Carlo output, i.e. only amino acid substitutions which occur in 10 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. This combinatorial library has a complexity of 336.

25

#### Interface Stability

The stability of an antibody can be increased by designing more favorable interactions between individual Ig domains at inter-Ig domain interfaces. For example, as can be seen in Figure 1, for human antibodies there are five interdomain interfaces that can be optimized using computational

30 screening methodology:  $V_H/V_L$ ,  $C\gamma 1/C_L$ ,  $V_H/C\gamma 1$ ,  $V_L/C_L$ , and  $C\gamma 3/C\gamma 3$ .

#### *Example 6: rhumAb VEGF $V_H/V_L$ Interface Stabilization*

The stability of the interface between the  $V_H$  and  $V_L$  domains is critical to antibody stability. The antibody rhumAb VEGF was stabilized by enhancing the interaction between the  $V_H$  and  $V_L$  domains

35 by using computational screening methods to design more favorable interactions between the residues which make up this interface. rhumAb VEGF is a humanized antibody that is currently in clinical development for treatment of a variety of cancers. The high resolution structure is available of the complex of the rhumAb VEGF Fab fragment with its target antigen, the vascular endothelial growth factor (VEGF). This structure, PDB accession code 1CZ8, served as the template for design

40 calculations. The  $V_H/V_L$  interface of rhumAb VEGF is shown in Figure 14. Variable positions were chosen by visual inspection of the 1CZ8 structure, and these positions are shown in Figure 14 and listed in Figures 15a and 15b. For rhumAb VEGF, the interface can be separated into two somewhat independent sets of residues, and thus it was possible to carry out computational screening in two separate sets of design calculations. The sets of amino acids considered at variable positions were

5 chosen subjectively by visual inspection of the 1CZ8 structure. The conformations of amino acids at variable positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library.

10 The 1CZ8 structure was used as the template for design calculations. For both sets of calculations, the energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequences were determined using a DEE algorithm. These ground states, and the WT rhumAb VEGF sequence, are shown in Figures 15a and 15b. The fact that the predicted ground state sequences are very similar to  
15 the WT sequence validates the computational screening method. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground states. Figures 15a and 15b show the output sequence lists from these Monte Carlo searches.

20 These results can be used to generate one or more experimental libraries which can be screened for increased antibody stability. An experimental library, shown in Figures 15c, was derived by applying a 1% cutoff of occupancy to the Monte Carlo output from each set of calculations, and then these primary libraries were subsequently combined to generate a secondary library with mutations at all positions. This combinatorial library has a complexity of  $1.3 \times 10^7$ .

25 Because of the number of residues involved in mediating this interface, it may be beneficial to reduce the complexity of the design calculations. As discussed above, sequence information can be used to guide the choice of variable positions and the set of amino acids considered at those positions. The use of sequence information here will enable the complexity of the computational problem to be  
30 reduced while ensuring that the remaining diversity sampled is of high quality, in terms of the structural, functional, and immunogenic fidelity of the antibody. Figures 16a and 16b show the 1CZ8 heavy and light chain variable chain sequences aligned with the human  $V_H$  and  $V_L$  kappa germ line sequences. A new design calculation using this information was run to stabilize the  $V_H/V_L$  interface. The sequence information was first used to reevaluate the list of variable positions. A subset of the  
35 positions in Figures 15a and 15b were chosen based on the degree of variability at each position in the germ line. Those positions with one type of amino acid in the majority of the sequences, or for which there is no sequence information, were not allowed to vary in the calculation. This new set is shown in Figure 17a. Light chain position 98 and heavy chain positions 45, 110, and 113 were not variable positions in this calculation, but were floated. The sequence information was also used to  
40 choose the set of amino acids to be considered at variable positions in the new design calculation. All amino acids, and only those amino acids, which appear at each variable position in the germ line were considered in the new design calculation. For variable positions in the light and heavy chain CDR3s, for which no sequence information is available, all 20 amino acids were considered. This set of considered amino acids is shown in Figure 17a.

5

The 1CZ8 structure was used as the template for design calculations. In this new calculation, energies of all possible combinations were not precalculated. Instead, a genetic algorithm was used to screen for low energy sequences, with energies being calculated during each round of "evolution" only for those sequences being sampled. The conformations of amino acids at variable and floated positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library using a flexible rotamer model (Mendes *et. al.*, 1999, *Proteins: Structure, Function, and Genetics* 37:530-543). Energies were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions. This calculation generated a list of 300 sequences which are predicted to be low in energy. Clustering was performed to facilitate analysis of the results and library generation. The 300 output sequences were clustered computationally into 10 groups of similar sequences using a nearest neighbor single linkage hierarchical clustering algorithm to assign sequences to related groups based on similarity scores (Diamond, R., Coordinate-Based Cluster Analysis, *Acta Cryst.* 1995 , D51, 127-135.). That is, all sequences within a group are most similar to all other sequences within the same group and less similar to sequences in other groups. The lowest energy sequence from each of these ten clusters, used here as a representative of each group, is presented in Figure 17a.

These results can be used to generate one or more experimental libraries which can be subsequently screened for increased antibody stability. An experimental library can be derived directly from the representative cluster group sequences. Thus Figure 17a provides a 10 sequence experimental library. To efficiently use experimental resources, this library size of 10 variants could be screened first, followed by subsequent screening of sequences or a subset of sequences within the group to which the experimentally determined most favorable variant belongs. For example, if variant 5 (i.e. the lowest energy sequences from cluster group 5) was found to be most favorable, all of the sequences of cluster group 5 could be subsequently screened. The 14 sequences in group 5 are presented in Figure 17b as an example of such an experimental library.

#### *Example 7: Herceptin V<sub>H</sub>/V<sub>L</sub> Interface Stabilization*

The interface between the V<sub>H</sub> and V<sub>L</sub> domains of the antibody Herceptin was also stabilized. More favorable interactions between the V<sub>H</sub> and V<sub>L</sub> domains were designed using computational screening methods. Herceptin, which targets the extracellular domain of the proto-oncogene Her2/neu gene product, also known as erbB2, is a humanized antibody that is currently marketed for treatment for breast cancer. The high resolution structure is available of uncomplexed Herceptin scFv. This structure, PDB accession code 1FVC, served as the template for design calculations. The V<sub>H</sub>/V<sub>L</sub> interface of Herceptin is shown in Figure 18. Variable positions were chosen by visual inspection of the 1FVC structure, and these positions are shown in Figure 18 and listed in Figure 19a. The majority of the chosen core variable positions are sequestered from solvent, and therefore the amino acids conserved were chosen as the set belonging to the core classification. The exception is light chain position 43, substitutions at which could potentially make favorable polar interactions, and so amino

5 acids considered for this position were chosen as the set belonging to the boundary classification. The conformations of amino acids at variable positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library.

10 The 1FVC structure was used as the structural template for design calculations. The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This ground state, and the WT Herceptin sequence, are shown in Figure 19a. The fact that the  
15 predicted ground state sequence is very similar to the WT sequence validates the computational screening technology. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 19a shows the output sequence list from this Monte Carlo search. These results can be used to generate one or more experimental libraries which can be subsequently screened for increased antibody stability. An experimental library, shown in Figure 19b, was derived  
20 by applying a 1% cutoff of occupancy to the Monte Carlo output from each set of calculations, i.e. only amino acid substitutions which occur in 10 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. Additionally, the glutamine was added at light chain position 89 so that the WT sequence is represented. This combinatorial library has a complexity of 5184. In the above calculation, for all but one variable position only nonpolar amino acids were considered.  
25 As discussed above, nonpolar residues have a higher tendency to aggregate than polar residues, and therefore nonpolar amino acids at the interdomain interface can result in a greater nonreversibility of the unbinding/binding transition. Design of a stable interface with greater polar character may thus provide greater thermodynamic reversibility and improved solubility. Another Herceptin V<sub>H</sub>/V<sub>L</sub> interface calculation was carried out in which the amino acids considered were chosen as the set  
30 belonging to the surface classification. A number of nonpolar interactions, however, appear critical to this interface, both by visual inspection and by their level of conservation in the aligned germ lines (Figures 2a and 2b). These positions, including light chain positions 36 and 89, and heavy chain positions 95 and 110, were floated in the new calculation. The remaining set of variable positions is shown in Figure 19c.

35 The 1FVC structure was used as the template for design calculations. The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This  
40 ground state, and the WT Herceptin sequence, are shown in Figure 19c. The fact that the predicted ground state sequence is very similar to the WT sequence validates the computational screening technology. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 19c shows the output sequence list from this Monte Carlo search.

5

These results can be used to generate one or more experimental libraries which can be screened for increased antibody stability. An experimental library, shown in Figure 19d, was derived by applying a 5% cutoff of occupancy to the Monte Carlo output from each set of calculations, i.e. only amino acid substitutions which occur in 50 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. Additionally, the WT residues were added to the library so that the sequence space sampled experimentally also includes interfaces made up of favorable polar and nonpolar residues at these positions. This combinatorial library has a complexity of 4032.

*Example 8: rhumAb VEGF C<sub>L</sub>/C<sub>γ</sub>1 Interface Stabilization*

15 The interface between the C<sub>L</sub> and C<sub>γ</sub>1 domains can also be stabilized using computational screening. More favorable interactions were designed between residues which make up the rhumAb VEGF C<sub>L</sub>/C<sub>γ</sub>1 interface. The C<sub>L</sub>/C<sub>γ</sub>1 interface of rhumAb VEGF is shown in Figure E8. Variable positions were chosen by visual inspection of the 1CZ8 structure, and these positions are shown in Figure 20 and listed in Figure 21a. Because these positions are almost completely sequestered from solvent, 20 the amino acids considered were chosen as the set belonging to the core classification, even for 176, 178, and 189 which are polar amino acids in the WT sequence. The WT amino acids were, however, also considered at these positions. The conformations of amino acids at variable positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library. The 1CZ8 structure was used as the template for design calculations. The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a 25 force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This ground state, and the WT rhumAb VEGF sequence, are shown in Figure 21a. The fact that the predicted ground state sequence is very similar to the WT sequence validates the computational screening method. A diversity of sequences for an experimental library was generated by using a 30 Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 21a shows the output sequence list from this Monte Carlo search. These results can be used to generate one or more experimental libraries which can be subsequently screened for increased antibody stability. An experimental library, shown in Figure 21b, was derived 35 by applying a 5% cutoff of occupancy to the Monte Carlo output from each set of calculations, i.e. only amino acid substitutions which occur in 50 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. Three additional amino acids were added to this library: threonine and serine were added to light chain position 178 and heavy chain position 189 respectively so that all polar residues are represented in the library, and the valine at light chain position 178 was 40 also included even though it did not make the 5% cutoff. As is known in the art, valine is a good nonpolar substitution for threonine because the two have nearly identical size and shape. This combinatorial library has a complexity of 5184.

5     *Example 9: Fc C $\gamma$ 3/C $\gamma$ 3 Interface Stabilization*

The interface between the C $\gamma$ 3 domains can also be stabilized using computational screening. Again, because this domain is a part of the antibody Fc region, improvements made are widely applicable to antibodies, independent of what antigen is bound at the variable region. More favorable interactions were designed between residues which make up the Fc C $\gamma$ 3/C $\gamma$ 3 interface. Variable positions were  
10    chosen by visual inspection of the 1DN2 structure, and these positions are shown in Figure 22 and listed in Figure 23a. Because these positions are almost completely sequestered from solvent, the amino acids considered were chosen as the set belonging to the core classification, although the WT amino acid was included at each position. The conformations of amino acids at variable positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer  
15    library.

The 1DN2 structure was used as the template for design calculations. The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond  
20    interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This ground state, and the WT Fc sequence, are shown in Figure 23a. The fact that the predicted ground state sequence is very similar to the WT sequence validates the computational screening method. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 23a  
25    shows the output sequence list from this Monte Carlo search.

These results can be used to generate one or more experimental libraries which can be subsequently screened for increased antibody stability. An experimental library, shown in Figure 23b, was derived by applying a 5% cutoff of occupancy to the Monte Carlo output from each set of calculations, i.e. only  
30    amino acid substitutions which occur in 50 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. This combinatorial library has a complexity of 1800.

*Solubility Optimization*

As discussed above, computational screening methods can be used to optimize the solubility of  
35    antibodies by designing favorable, more soluble substitutions at surface exposed nonpolar residues. Residues which can be replaced include residues which are exposed to solvent on individual Ig domains, including V<sub>H</sub>, V<sub>L</sub>, C $\gamma$ 1, C<sub>L</sub>, C $\gamma$ 2, and C $\gamma$ 3 as well as the linkers and/or hinges that connect them, or which lie at the interface between Ig domains.

40    *Example 10: Campath Solubility Optimization*

All four Ig domains of the Campath Fab antibody fragment were optimized for greater solubility using computational screening. Computational screening was applied to evaluate the replacement of all exposed nonpolar residues on these domains, including V<sub>H</sub>, V<sub>L</sub>, C $\gamma$ 1, C<sub>L</sub>, with all 20 amino acids. Variable positions were chosen by visual inspection of the 1CE1 structure, and include exposed

5 nonpolar residues which are not involved in binding antigen. These positions are shown in Figure 24 and listed in Figure 25a. Each of the 20 amino acids was considered at each variable position. The 1CE1 structure was used as the template for design calculations. For each variable position, each of the 20 amino acids was substituted and allowed to sample rotamer conformations derived from a backbone-independent rotamer library using a flexible rotamer model. A genetic algorithm was  
10 used to optimize the conformation of each amino acid substitution at each variable position, with energies being calculated during each round of evolution. In this way, the lowest energy rotamer of each substitution was determined, and this energy was defined as the energy of substitution for that amino acid at that variable position. Thus this design calculation provided an energy of substitution for each of the 20 amino acids at each variable position. Figure 25a shows these results. At each  
15 variable position, the lowest energy substitution and all amino acid substitutions which are within 1 unit of energy of the lowest energy substitution are shown. Thus Figure 25a presents the most favorable substitutions for each of the variable positions.

These results can be used to generate one or more experimental libraries which can be subsequently  
20 screened for improved antibody solubility. An experimental library was derived from this computational screening output by including the WT amino acid and all favorable polar amino acid substitutions at each variable position. As can be seen, no polar substitutions are predicted to be favorable for heavy chain position 116, and so this position is left as the WT leucine in the library. This experimental library, which has a combinatorial complexity of 11200, is shown in Figure 25b.

25

*Example 11: rhumAb VEGF Solubility Optimization*

All four Ig domains of the rhumAb VEGF Fab antibody fragment were optimized for greater solubility using computational screening. Computational screening was applied to evaluate the replacement of all exposed nonpolar residues on these domains, including V<sub>H</sub>, V<sub>L</sub>, C<sub>γ</sub>1, C<sub>L</sub>, with all 20 amino acids.  
30 Variable positions were chosen by visual inspection of the 1CZ8 structure, and include exposed nonpolar residues which are not involved in binding antigen. These positions are shown in Figure 26 and listed in Figure 27a. Each of the 20 amino acids was considered at each variable position. The 1CZ8 structure was used as the template for design calculations. For each variable position, each of the 20 amino acids was substituted and allowed to sample rotamer conformations derived  
35 from a backbone-independent rotamer library using a flexible rotamer model. A genetic algorithm was used to optimize the conformation of each amino acid substitution at each variable position, with energies being calculated during each round of evolution using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions. In this way, the lowest energy rotamer of each substitution was determined. This energy was defined as the energy of  
40 substitution for that amino acid at that variable position. Thus this design calculation provided an energy of substitution for each of the 20 amino acids at each variable position. Figure 27a shows these results. At each variable position, the lowest energy substitution and all amino acid substitutions which are within 1 unit of energy of the lowest energy substitution are shown. Thus Figure 27a presents the most favorable substitutions for each of the variable positions.

5

These results can be used to generate one or more experimental libraries which can be subsequently screened for improved antibody solubility. An experimental library was derived from this computational screening output by including the WT amino acid and all favorable polar amino acid substitutions at each variable position. As can be seen, no polar substitutions are predicted to be favorable for light chain positions 15 and 125 and heavy chain positions 80, 118, and 169, and so these positions are left as the nonpolar WT amino acids in the library. This experimental library, which has a combinatorial complexity of 61440, is shown in Figure 27b.

*Example 12: Herceptin Solubility Optimization*

As discussed above, by removing certain regions or domains of an antibody to generate an antibody fragment, nonpolar residues that make up the interface with another Ig domain in the context of a full-length antibody or larger antibody fragment can become exposed. For example, for Herceptin, the  $V_H$  and  $V_L$  residues which make up the  $V_H/C\gamma 1$  and  $V_L/C_L$  interfaces are exposed to solvent in an scFv fragment, as is seen in the 1FVC structure. Computational screening was used to engineer favorable, more soluble mutations at these positions for Herceptin. Variable positions were chosen by visual inspection of the 1FVC structure, and include the set of exposed nonpolar residues at the C-terminal end of the  $V_H$  and  $V_L$  domains. These positions are shown in Figure 28 and listed in Figure 29a. Each of the 20 amino acids was considered at each variable position.

The 1FVC structure was used as the template for design calculations. For each variable position, each of the 20 amino acids was substituted and allowed to sample rotamer conformations derived from a backbone-independent rotamer library using a flexible rotamer model. A genetic algorithm was used to optimize the conformation of each amino acid substitution at each variable position, with energies being calculated during each round of evolution using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions. In this way, the lowest energy rotamer of each substitution was determined, and this energy was defined as the energy of substitution for that amino acid at that variable position. Thus this design calculation provided an energy of substitution for each of the 20 amino acids at each variable position. Figure 29a shows these results. At each variable position, the lowest energy substitution and all amino acid substitutions which are within 1 unit of energy of the lowest energy substitution are shown. Thus Figure 29a presents the most favorable substitutions for each of the variable positions.

These results can be used to generate one or more experimental libraries which can be subsequently screened for improved antibody solubility. An experimental library was derived from this computational screening output by including the WT amino acid and all favorable polar amino acid substitutions at each variable position. As can be seen, no polar substitutions are predicted to be favorable for light chain position 83, and so this position is left as the nonpolar WT phenylalanine in the library. This experimental library, which has a combinatorial complexity of 2530, is shown in Figure 29b.

5

*Example 13: Fc Solubility Optimization*

The Fc region was optimized for greater solubility using computational screening. Computational screening was applied to evaluate the replacement of all exposed nonpolar residues on the C $\gamma$ 2 and C $\gamma$ 3 domains with all 20 amino acids. Variable positions were chosen by visual inspection of the 1DN2 structure, and include exposed nonpolar residues which are not involved in binding an Fc receptor. For example Met252 and Met428 are involved in binding to FcRn (Martin *et al.*, 2001, *Mol. Cell* 7:867-877), and Tyr296 and Tyr300 are close to the binding site for Fc $\gamma$ Rs (Sonderman *et al.*, 2001, *J. Mol. Biol.* 309:737-749). Therefore these residues, despite being exposed nonpolar, were not included as variable positions. Variable positions are shown in Figure 30 and listed in Figure 31a. Each of the 20 amino acids was considered at each variable position.

The 1DN2 structure was used as the template for design calculations. For each variable position, each of the 20 amino acids was substituted and allowed to sample rotamer conformations derived from a backbone-independent rotamer library using a flexible rotamer model. A genetic algorithm was used to optimize the conformation of each amino acid substitution at each variable position, with energies being calculated during each round of evolution using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions. In this way, the lowest energy rotamer of each substitution was determined. This energy was defined as the energy of substitution for that amino acid at that variable position. Thus this design calculation provided an energy of substitution for each of the 20 amino acids at each variable position. Figure 31a shows these results. At each variable position, the lowest energy substitution and all amino acid substitutions which are within 1 unit of energy of the lowest energy substitution are shown. Thus Figure 31a presents the most favorable substitutions for each of the variable positions. These results can be used to generate one or more experimental libraries which can be subsequently screened for improved antibody solubility. An experimental library was derived from this computational screening output by including the WT amino acid and all favorable polar amino acid substitutions at each variable position. As can be seen, no polar substitutions are predicted to be favorable for position 404, and so this position was left as the nonpolar WT phenylalanine in the library. This experimental library, which has a combinatorial complexity of  $4.9 \times 10^8$ , is shown in Figure 31b.

*Affinity Maturation*

As discussed above, a number of strategies can be applied for utilizing computational screening methodology to affinity mature antibodies.

40

*Example 14: rhumAb VEGF Affinity Maturation Using The Antibody/Antigen Complex Structure*

The availability of the bound antibody/antigen structure for rhumAb VEGF enables the affinity of this antibody to be enhanced directly using computational screening. More favorable interactions between the rhumAb VEGF antibody and its antigen were designed. Variable positions involved in mediating

5 this interaction were chosen by visual inspection of the 1CZ8 structure, shown in Figure 32 and listed in Figure 33a. The set of amino acids allowed at variable positions was also chosen by visual inspection. Antigen residues which contact variable residue positions were floated. The conformations of amino acids at variable and floated positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library.

10 The 1CZ8 structure was used as the template for design calculations. The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This  
15 ground state, and the WT rhumAb VEGF sequence, are shown in Figure 33a. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 33a shows the output sequence list from this Monte Carlo search.

20 These results can be used to generate one or more experimental libraries which can be screened for enhanced affinity for antigen. An experimental library, shown in Figure 33b, was derived by applying a 5% cutoff of occupancy to the Monte Carlo output from each set of calculations, i.e. only amino acid substitutions which occur in 50 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. Additionally, the WT amino acids at heavy chain positions 31, 54, 57, and 59  
25 were added to the library so that the WT sequence is represented combinatorially in the library. This experimental library has a complexity of 2304.

In another set of calculations, rhumAb VEGF was affinity matured by reengineering antibody residues which do not contact antigen. Here the variable positions in the design calculation were those  
30 residues which interact with contact residues, but are not themselves contact residues. As discussed above, by using computational screening to explore substitutions in the shell of residues which interact with contact residues, a quality diversity of new contact residue conformations can be sampled. Variable positions involved were chosen by visual inspection of the 1CZ8 structure, shown in Figure 34 and listed in Figure 35a. The set of amino acids allowed at variable positions was also  
35 chosen by visual inspection. The conformations of amino acids at variable positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library. The 1CZ8 structure was used as the template for design calculations. The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond  
40 interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This ground state, and the WT rhumAb VEGF sequence, are shown in Figure 35a. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 35a shows the output sequence list from this Monte Carlo search.

5

These results can be used to generate one or more experimental libraries which can be screened for enhanced affinity for antigen. An experimental library, shown in Figure 35b, was derived by applying a 5% cutoff of occupancy to the Monte Carlo output from each set of calculations, i.e. only amino acid substitutions which occur in 50 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. The WT is already represented in this library, and so no additional amino acids were added. This experimental library has a complexity of 784.

*Example 15: SM3 Affinity Maturation Using The Antibody/Antigen Complex Structure*

The availability of the bound antibody/antigen complex structure for SM3 enables the affinity of this antibody to be enhanced directly using computational screening. SM3 is a mouse antibody that is currently being developed as an anticancer agent. The high resolution structure is available of the complex of the SM3 Fab with its target antigen, a peptide from the cell surface mucin MUC1. This structure, PDB accession code 1SM3, served as the template for design calculations. More favorable interactions between the SM3 antibody and its antigen were designed. SM3 binds the MUC1 peptide using an extensive binding pocket which involves a large number of SM3 residues. The pocket can, however, be separated into two somewhat independent sets of residues, and thus in order to reduce the complexity of the computational screen, two separate sets of design calculations were carried out. Variable positions involved in mediating this interaction were chosen by visual inspection of the 1SM3 structure, shown in Figure 36 and listed in Figure 37a and 37b. The set of amino acids allowed at variable positions was also chosen by visual inspection. Antigen residues were kept fixed in the two calculations. The conformations of amino acids at variable positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library.

The 1SM3 structure was used as the template for design calculations. For both sets of calculations, the energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequences were determined using a DEE algorithm. These ground states, and the WT SM3 sequence, are shown in Figure 37a and 37b. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground states. Figure 37a and 37b show the output sequence lists from these Monte Carlo searches. These results can be used to generate one or more experimental libraries which can be subsequently screened for enhanced affinity for antigen. An experimental library, shown in Figure 37c, was derived by applying a 5% cutoff of occupancy to the Monte Carlo output from each set of calculations, and then these primary libraries were subsequently combined to generate a secondary library with mutations at all positions. Additionally, the WT amino acids at light chain positions 50, 53, 56, and 93, and heavy chain position 96 were added to the library so that the WT sequence is represented combinatorially in the library. This may be particularly important here because some glycine and proline residues in the WT sequence were allowed to be variable in the calculations. These amino

5 acids can be important determinants of protein backbone conformation, and therefore the benefit of their replacement with side chains which are capable of making favorable interaction with antigen may be outweighed by unfavorable potential backbone movements. This combinatorial library has a complexity of  $3.5 \times 10^6$ .

10 *Example 16: Campath Affinity Maturation Using The Antibody/Antigen Complex Structure*

The availability of the bound antibody/antigen complex structure for Campath enables the affinity of this antibody to be enhanced directly using computational screening. More favorable interactions between the Campath antibody and its antigen were designed. Variable positions involved in mediating this interaction were chosen by visual inspection of the 1CE1 structure, shown in Figure 38 and listed in Figure 39a. The set of amino acids allowed at variable positions was also chosen  
15 subjectively by visual inspection. Antigen residues were floated. The conformations of amino acids at variable and floated positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library.

20 The 1CE1 structure was used as the template for design calculations. The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This ground state and the WT Campath sequence are shown in Figure 39a. A diversity of sequences for  
25 an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground state. Figure 39a shows the output sequence list from this Monte Carlo search.

These results can be used to generate one or more experimental libraries which can be screened for  
30 enhanced affinity for antigen. An experimental library, shown in Figure 39b, was derived by applying a 5% cutoff of occupancy to the Monte Carlo output from each set of calculations, i.e. only amino acid substitutions which occur in 50 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. Additionally, the WT asparagine at light chain position 50 was added to the library so that the WT sequence is represented combinatorially in the library. This combinatorial  
35 library has a complexity of 486.

Because of the number of residues involved in mediating the interaction of Campath with its antigen, it may be beneficial to reduce the complexity of the design calculations. The use of sequence  
information here will enable the complexity of the computational problem to be reduced while ensuring  
40 that the remaining diversity sampled is of high quality, in terms of the structural, functional, and immunogenic fidelity of the antibody. Sequence information was used to guide the choice of variable positions and the set of amino acids considered at those positions for the Campath affinity maturation calculations. Figures 40a and 40b show the Campath heavy and light chain variable chain sequences aligned with the human  $V_H$  and  $V_L$  kappa germ line sequences. A new design calculation using this

5 information was run to affinity mature Campath. The sequence information was first used to reevaluate the list of variable positions. A subset of the positions in Figure 39a was chosen based on the degree of variability at each position in the germ line. The sequence information was used to choose the set of amino acids considered at variable positions in the new design calculation. All amino acids, and only those amino acids, which appear at each variable position in the germ line were  
10 considered in the new design calculation. For variable positions in CDR3, for which no sequence information is available, all 20 amino acids were considered. This set of amino acids is shown in Figure 41a. Antigen residues were allowed to float during the calculations.

The 1CE1 structure was used as the template for design calculations. In this new calculation,  
15 energies of all possible combinations were not precalculated. Instead, a genetic algorithm was used to screen for low energy sequences, with energies being calculated during each round of "evolution" only for those sequences being sampled. The conformations of amino acids at variable and floated positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library using a flexible rotamer model. Energies were calculated using a force field containing  
20 terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions. This calculation generated a list of 300 sequences which are predicted to be low in energy. Clustering was performed to facilitate analysis of the results and library generation. The 300 output sequences were clustered computationally into 10 groups of similar sequences using a nearest neighbor single linkage hierarchical clustering algorithm to assign sequences to related groups based on similarity scores  
25 (Diamond, R., Coordinate-Based Cluster Analysis, Acta Cryst. 1995 , D51, 127-135.). The 300 output sequences were clustered computationally into 10 groups of similar sequences. That is, all sequences within a group are most similar to all other sequences within the same group and less similar to sequences in other groups. The lowest energy sequence from each of these ten clusters, used here as a representative of each group, is presented in Figure 41a.

30 These results can be used to generate one or more experimental libraries which can be subsequently screened for increased affinity for antigen. An experimental library can be derived directly from the representative cluster group sequences. Thus Figure 41a provides a 10 sequence experimental library. To efficiently use experimental resources, this library size of 10 variants could be screened  
35 first, followed by subsequent screening of sequences or a subset of sequences within the group to which the experimentally determined most favorable variant belongs. For example, if variants 4 and 9 (i.e. the lowest energy sequences from cluster groups 4 and 9) were found experimentally to be most favorable, all of the sequences of cluster groups 4 and 9 could be subsequently screened. The 6 sequences in group 4 and 5 sequences in group 9 are presented in Figure 41b as an example of such  
40 an experimental library.

#### *Example 17: D3H44 Affinity Maturation Using Complex And Uncomplexed Structures*

The availability of structural information for both the bound and unbound forms of the anti-tissue factor antibody D3H44 provide the opportunity to explore how both complexed and uncomplexed structural

5 information can be used to computationally affinity mature an antibody. D3H44 is a humanized antibody that is currently being developed for treatment of thrombotic disorders. The high resolution structure of the D3H44 antibody/antigen complex, PDB accession code 1JPT, and the unbound antibody structure, PDB accession code 1JPS, served as templates in separate sets of design  
10 calculations aimed at designing more favorable interactions between the D3H44 antibody and its antigen. Variable positions involved in mediating this interaction were chosen by visual inspection of the 1JPT structure, shown in Figure 42 and listed in Figure 43a. The set of amino acids considered at variable positions was also chosen by visual inspection. Antigen residues which contact antibody  
15 variable position residues were floated in the bound structure calculation. The conformations of amino acids at variable and floated positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library.

The 1JPT and 1JPS structures were used as templates in two separate sets of design calculations. For both sets of calculations, the energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van  
20 der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequences were determined using a DEE algorithm. These ground states, and the WT D3H44 sequence, are shown in Figures 43a and 43b. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of 1000 similar sequences around the predicted ground states. Figures 43a and 43b show the output sequence lists from these  
25 Monte Carlo searches.

Notably, the diversity of sequences in the bound output is approximately a subset of the sequences in the unbound output. This result validates the use of using unbound structural information for affinity maturation, because it indicates that such calculations, while reducing sequence complexity for  
30 experimental screening, still produce quality antigen binding diversity. That is, experimental libraries derived from such calculations are enriched in sequences that favorably bind antigen. For example, experimental libraries were generated from the output of both bound and unbound calculations. These experimental libraries, shown in Figure 43c, were derived by applying a 1% cutoff of occupancy to the Monte Carlo output from each set of calculations, i.e. only amino acid substitutions which occur  
35 in 10 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. Additionally, WT amino acids were incorporated into the library if they were not already represented. The combinatorial complexities are 1296 and 211680 for the bound- and unbound-derived libraries respectively. As can be seen, a significant portion of the sequences present in the bound-derived library are present in the unbound-derived library, which is substantially reduced in complexity from  
40 random sequences.

The results from both sets of calculations can be combined to generate an experimental library. An experimental library, shown in Figure 43d, was derived by including only those substitutions which are present in the Monte Carlo outputs of both bound and unbound design calculations. Additionally, the

5 WT amino acid at light chain position 94 was added to the library so all of the WT amino acids are represented. This library provides a list of substitutions that are compatible with the antibody in both forms, ensuring that the derived library does not contain variants that are poorly behaved in the absence of antigen. Furthermore, substitutions which are favorable in the bound form but unfavorable in the unbound form may be due to the need for significant conformational changes for binding.

10 Elimination of these substitutions may trim the library of unfavorable variants which lose entropy upon binding. This combinatorial library has a complexity of 864.

*Example 18: Herceptin Affinity Maturation Using The Uncomplexed Structure*

Although there is a structure available of the unbound Herceptin scFv antibody fragment, there is no available structure of the bound antibody/antigen complex. However, there is a wealth of experimental information available which can be used to guide affinity maturation design calculations.

15 An alanine scanning mutagenesis study (Kelley *et al.*, 1993, *Biochemistry* 32:6828-6835) showed that there are four central Herceptin residues, W, X, Y, and Z which are crucial for binding the Her2/neu antigen. A subsequent study used phage display to explore sequence diversity at these residues and residues proximal to them in the 1FVC structure (Gerstner *et al.*, 2002, *J. Mol. Biol.* 321:851-862).

20 The results from these studies were used to guide the choice of variable positions and amino acids considered at those positions in design calculations aimed at affinity maturing the Herceptin antibody. Here the goal is to utilize computational screening to generate a high quality library that is enriched for substitutions at antigen binding positions which are structurally compatible with the Herceptin antibody.

25 Variable positions were chosen as those positions which show moderate variability in the phage display results. That is, positions that were very intolerant to mutation (one amino acid identity was observed in the majority of selected sequences), and positions that were very tolerant to mutation (no preference for amino acid identity was observed) were not chosen as variable positions. Mutations at these positions are expected to have a deleterious effect or no effect respectively on antigen binding.

30 Positions that have some but not stringent amino acid requirements have the most value in terms of exploring diversity which may be more favorable for antigen binding. These positions are shown in Figure 44 and listed in Figure 45a. The set of amino acids considered at these variable positions was also guided by the experimental results. For a given position, if the diversity of substitutions observed was greater than 90% polar or nonpolar residues, the amino acids considered for that position were chosen as the set belonging to the surface or core classification respectively. If no trend was observed, the amino acids considered for that position were chosen as the set belonging to the boundary classification. The conformations of amino acids at variable positions were represented as a set of side chain rotamers derived from a backbone-independent rotamer library.

35 The 1FVC structure was used as the template for design calculations. The energies of all possible combinations of the considered amino acids at the chosen variable positions were calculated using a force field containing terms describing van der Waals, solvation, electrostatic, and hydrogen bond interactions, and the optimal (ground state) sequence was determined using a DEE algorithm. This ground state, and the WT Herceptin sequence, are shown in Figure 45a.. A diversity of sequences for an experimental library was generated by using a Monte Carlo algorithm to evaluate the energies of

40

5     1000 similar sequences around the predicted ground state. Figure 45a. shows the output sequence list from this Monte Carlo search.

10     These results can be used to generate one or more experimental libraries which can be screened for enhanced affinity for antigen. An experimental library, shown in Figure 45b, was derived by applying a 1% cutoff of occupancy to the Monte Carlo output from each set of calculations, i.e. only amino acid substitutions which occur in 10 or greater variants out of the 1000 Monte Carlo output sequences are included in the library. Additionally, the WT amino acids at light chain positions 53 and 91, and heavy chain positions 59 were added to the library so that the WT sequence is represented combinatorially in the library. This experimental library has a complexity of 16800.

15

All references cited herein are incorporated by reference in their entirety.

20     Whereas particular embodiments of the invention have been described above for purposes of illustration, it will be appreciated by those skilled in the art that numerous variations of the details may be made without departing from the invention as described in the appended claims.

5

## CLAIMS

We claim:

1. A method for optimizing at least one physico-chemical property of an antibody, said method executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of:
  - a. receiving a template antibody structure;
  - b. selecting at least one variable positions which belong to said template antibody structure;
  - c. selecting at least one amino acids to be considered at said variable positions;
  - d. analyzing the interaction of each of said amino acids at each variable position with at least part of the remainder of said antibody, including said amino acids at other variable positions; and
  - e. identifying a set of at least one antibody sequence with at least one optimized physico-chemical property.
2. A method according to claim 1, wherein at least one of the optimized physico-chemical properties is selected from the group consisting of stability, solubility, and antigen binding affinity.
3. A method according to claim 2, wherein at least one of the optimized physico-chemical properties is stability.
4. A method according to claim 3, wherein the stabilized portion of said antibody is selected from the group consisting of a domain and an interface between domains.
5. A method according to claim 4, wherein the stabilized portion of said antibody is a domain.
6. A method according to claim 4, wherein the stabilized portion of said antibody is an interface between domains.
7. A method according to claim 2, wherein the physico-chemical property is solubility.
8. A method according to claim 7, wherein at least one antibody sequence possesses an increase in polar character.
9. A method according to claim 7, wherein said selecting step further comprises selecting at least one nonpolar amino acid and substituting said nonpolar amino acid with a polar amino acid.

5

10. A method according to claim 7, wherein said selecting step further comprises altering the pI of the antibody.

11. A method according to claim 2, wherein at least one of the optimized physico-chemical properties is antigen binding affinity.

10

12. A method according to claim 11, wherein at least one of said variable positions is located in a framework region of the antibody.

15

13. A method according to claim 11, wherein at least one of said variable positions is located in a complementarity determining region (CDR) of the antibody.

14. A method according to claim 1, wherein each of said amino acids at each of said variable positions are represented as a group of potential rotamers.

20

15. A method according to claim 1, wherein at least two variable positions are selected and at least two amino acids are considered at each variable position.

25

16. A method according to claim 1, wherein said analyzing step further comprises a computational step utilizing at least two of the energy terms selected from the group consisting of van der Waals, electrostatics, hydrogen bonds and solvation.

17. A method according to claim 1, wherein said variable positions are chosen based on their level of variability in a set of aligned antibody sequences.

30

18. A method according to claim 1, wherein one said amino acids are chosen from a list of amino acids which occur at said position or positions in a set of aligned antibody sequences.

19. A method according to claim 1, wherein said analyzing step includes a Protein Design Automation program.

35

5        20. A method according to claim 1, wherein said analyzing step includes a Sequence Prediction Algorithm program.

21. A method according to claim 1, wherein said antibody is selected from the group consisting of a full-length antibody and an antibody fragment.

10

22. A method according to claim 1, wherein said antibody sequence is substantially encoded by at least one mammalian antibody gene.

23. A method according to claim 1, wherein said antibody is selected from the group consisting of  
15 a fully human antibody, a humanized antibody, a chimeric antibody, and an engineered antibody.

24. A method according to claim 1, further comprising f) generating a library from said set of at least one antibody sequence.

20        25. A method according to claim 24 wherein said library is a computational library.

26. A method according to claim 24 wherein said library is generated experimentally.

27. A method according to claim 24 further comprising g) experimentally screening said library.

25

28. A method according to claim 27, wherein said library is screened using at least one selection method.

29. A method according to claim 25, wherein said library is screened using at least one selection  
30 method selected from the group consisting of: phage display methods, cell surface display, in vitro display, and cytometric screening.

30. A method according to claim 25, wherein said selection method is a directed evolution method.

5        31. An antibody sequence from said library of claim 24.

32. An antibody sequence according to claim 28, wherein said antibody sequence is substantially encoded by a mammalian antibody gene.

10       33. An antibody identified from said screening of claim 24.

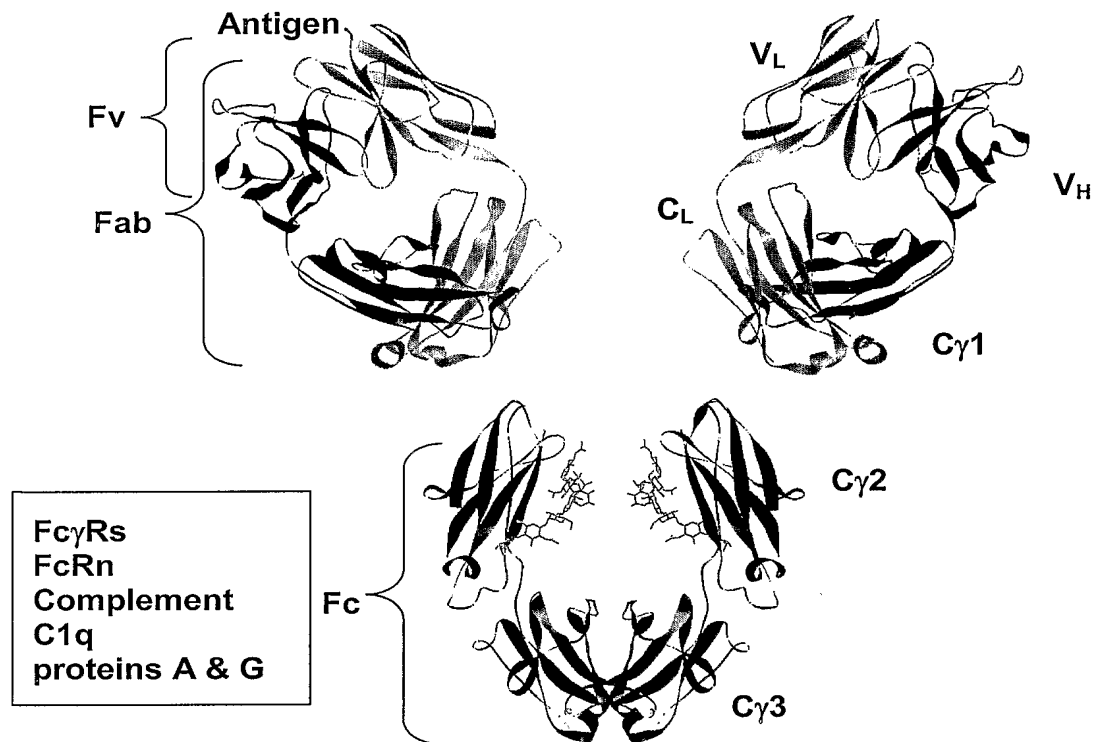
34. An antibody to claim according to claim 33, wherein said antibody is a full-length antibody or an antibody fragment.

15       35. An antibody according to claim 33, wherein said antibody is selected from the group consisting of a fully human antibody, a humanized antibody, a chimeric antibody, and an engineered antibody.

36. A method of treating a patient in need of said treatment, comprising administering an antibody of claim 28 to said patient.

20

Figure 1



[illegible]



**Figure 2b**

[illegible]

[illegible]

**Figure 3**

Antibody	Target	Source	PDB	Structure	Resolution (Å)	Reference
Campath	CD52	Humanized	1CE1	Bound	1.9	James <i>et al.</i> , 1999, <i>J. Mol. Biol.</i> 289:293-301
SM3	MUC1	Mouse	1SM3	Bound	2.0	Dokurno <i>et al.</i> , 1998, <i>J. Mol. Biol.</i> 284:713-728
D3H44	Tissue Factor	Humanized	1JPS	Bound	1.9	Faebler <i>et al.</i> , 2001, <i>J. Mol. Biol.</i> 313:83-97
D3H44	Tissue Factor	Humanized	1JPT	Unbound	1.9	Febler <i>et al.</i> , 2001, <i>J. Mol. Biol.</i> 313:83-97
Herceptin	Her2/Neu	Humanized	1FVC	Unbound	2.2	Eignebrat <i>et al.</i> , 1993, <i>J. Mol. Biol.</i> 229:969-995
rhumAb VEGF	VEGF	Humanized	1CZ8	Bound	2.4	Chen <i>et al.</i> , 1999, <i>J. Mol. Biol.</i> 293:865-881
Fc	-	Human IgG1	1DN2	Bound	2.7	Delano <i>et al.</i> , 2000, <i>Science</i> 287:1279-1283

Figure 4

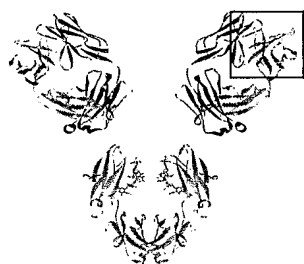


Figure 5a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
12 H	Core	V	I	I:701 V:299
18 H	Core	L	L	L:508 M:492
20 H	Core	L	L	L:966 M:34
36 H	Core	W	W	W:989 F:11
48 H	Core	I	M	I:698 M:301 V:1
66 H	Core	V	I	V:577 I:388 M:35
70 H	Core	V	V	V:1000
72 H	Core	M	M	M:1000
83 H	Core	L	L	L:1000
85 H	Core	L	L	M:490 L:475 I:29 V:6
88 H	Core	V	V	V:1000
117 H	Core	V	M	M:989 W:10 F:1

Figure 5b

Position	Experimental Library
12 H	I
18 H	L
20 H	L
36 H	W
48 H	M
66 H	I
70 H	V
72 H	M
83 H	L
85 H	L
88 H	V
117 H	M
Complexity	1

Figure 5c

Position	Experimental Library
12 H	I V
18 H	L M
20 H	L M
36 H	W F
48 H	I M
66 H	V I M
70 H	V
72 H	M
83 H	L
85 H	M L I
88 H	V
117 H	M W V
Complexity	864

Figure 6

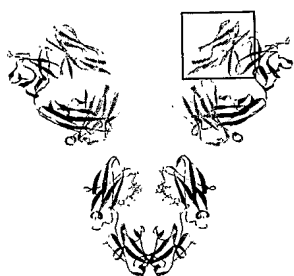


Figure 7a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
2 L	Core	I	I	I:935 V:54 L:11
4 L	Core	M	M	M:1000
29 L	Core	I	I	I:944 V:56
33 L	Core	L	L	L:1000
71 L	Core	F	F	F:996 M:4
92 L	Boundary	I	I	I:232 R:192 K:120 Y:100 V:97 M:91 F:53 L:51 T:28 H:28 E:4 A:2 Q:2
97 L	Boundary	T	T	T:300 F:167 W:156 I:87 M:87 Y:63 H:53 K:20 R:19 Q:13 V:13 S:12 E:10

Figure 7b

Position	Experimental Library
2 L	I V
4 L	M
29 L	I V
33 L	L
71 L	F M
92 L	I R K Y V M F L
97 L	T F W I M Y H
Complexity	448

Figure 8

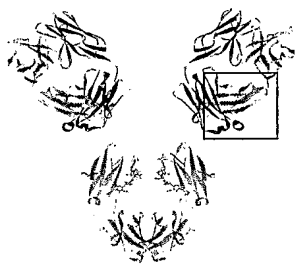


Figure 9a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
129 H	Core	V	V	V:1000
150 H	Core	V	M	M:656 V:287 I:51 L:6
153 H	Core	Y	F	F:771 Y:229
158 H	Core	V	V	V:1000
160 H	Core	V	V	V:881 I:119
173 H	Boundary	T	T	T:382 W:347 M:119 V:50 F:49 I:36 L:12 H:4 E:1
176 H	Core	A	A	A:1000
186 H	Core	L	L	L:767 M:233
188 H	Core	S	S	S:891 A:109
206 H	Core	V	V	V:1000

Figure 9b

Position	Experimental Library
129 H	V
150 H	M V I
153 H	F Y
158 H	V
160 H	V I
173 H	T W M V
176 H	A
186 H	L M
188 H	S A
206 H	V
Complexity	192

Figure 10

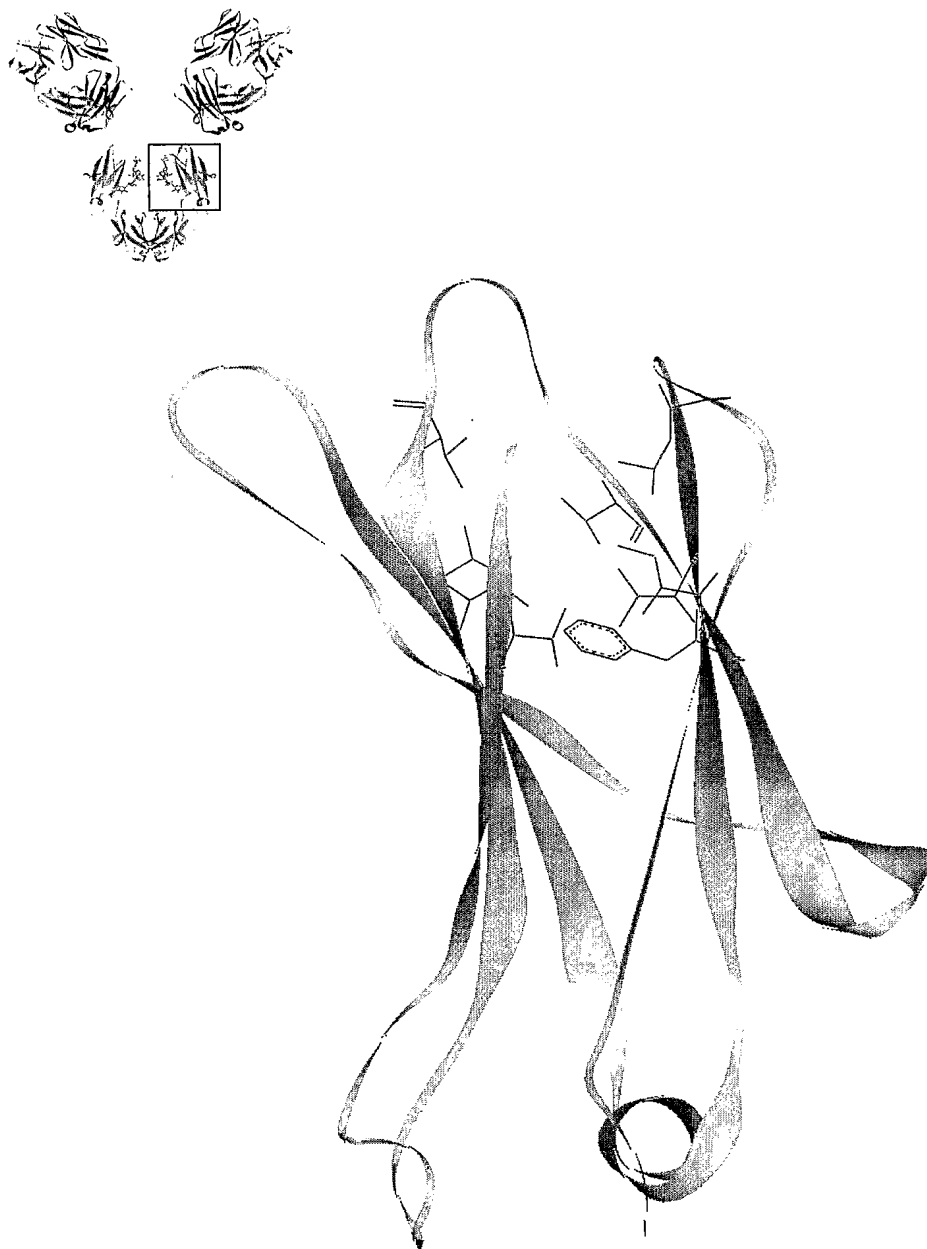


Figure 11a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
240	Core	V	V	V:981 M:19
263	Core	V	V	V:1000
266	Core	V	V	V:1000
273	Core	V	V	V:1000
275	Core	F	F	F:1000
302	Core	V	I	V:688 I:249 M:51 L:12
323	Core	V	V	V:913 I:68 M:19
328	Core	L	L	L:827 M:173
332	Boundary	I	I	I:440 M:182 W:116 E:74 V:72 L:59 K:57

Figure 11b

Position	Experimental Library
240	V M
263	V
266	V
273	V
275	F
302	V I M L
323	V I M
328	L M
332	I M W E V L K
Complexity	336

Figure 12

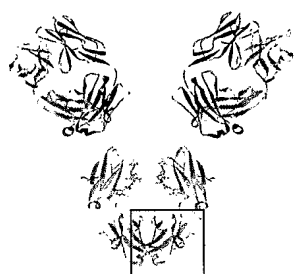


Figure 13a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
358	Boundary	M	M	M:626 L:100 T:84 W:82 E:42 K:37 I:18 S:6 Q:5
363	Core	V	V	V:993 L:7
365	Core	L	L	L:1000
381	Core	W	W	W:1000
391	Boundary	Y	F	Y:501 F:445 W:41 H:13
410	Core	L	L	L:1000
412	Core	V	V	V:987 I:13
417	Core	W	W	W:1000
423	Core	F	F	F:966 Y:34
441	Core	L	L	L:833 M:131 V:36

Figure 13b

Position	Experimental Library
358	M L T W E K I
363	V
365	L
381	W
391	Y F W H
410	L
412	V I
417	W
423	F Y
441	L M V
Complexity	336

Figure 14



Figure 15a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
36 L	Core	Y	F	Y:506 F:448 M:46
43 L	Boundary	A	Y	Y:493 F:292 H:179 A:20 N:8 S:8
87 L	Core	Y	F	F:606 Y:360 M:34
98 L	Core	F	F	F:989 M:11
35 H	Core	N	N	N:800 A:184 V:16
37 H	Core	V	V	V:945 I:55
45 H	Core	L	L	L:870 M:130
47 H	Core	W	W	W:1000
95 H	Core	Y	F	F:536 Y:391 M:61 L:12
110 H	Core	F	F	F:853 M:97 L:50
113 H	Core	W	W	W:1000

Figure 15b

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
32 L	Boundary	Y	Y	Y:603 M:137 F:95 E:47 K:42 L:28 Q:26 R:16 H:6
46 L	Core	V	V	V:1000
49 L	Boundary	Y	F	F:544 Y:432 M:24
50 L	Boundary	F	Y	Y:435 F:409 M:146 H:8 W:2
91 L	Core	Y	Y	Y:541 F:440 W:19
107 H	Boundary	H	F	F:691 M:190 H:115 Y:4
109 H	Core	Y	Y	Y:702 F:298

Figure 15c

Position	Experimental Library
32 L	Y M F E K L Q R
36 L	Y F M
43 L	Y F H A
46 L	V
87 L	F Y M
98 L	F M
35 H	N A V
37 H	V I
45 H	L M
47 H	W
49 L	F Y M
50 L	Y F M
91 L	Y F W
95 H	F Y M L
107 H	F M H
109 H	Y F
110 H	F M L
113 H	W
Complexity	13436928

**Figure 16a**

[illegible]



**Figure 16b**

[illegible]

[illegible]

Figure 17a

	Chain	L	L	L	L	L	L	L	L	H	H	H	H	H	H
	1CZ8 Position	32	36	43	46	49	50	87	91	35	37	47	95	107	109
	Chothia Position	32	36	43	46	49	50	87	91	35	37	47	91	100c	100e
	Amino Acids Considered	YW DA NS L	YF L	AV SP	VL SR HF	YF	YD AS YK LE TG WQ	Y F	YD LA FH TS GR	NH SQ G	VI AF	WY	YH C	All	All
	WT	Y	Y	A	V	Y	F	Y	Y	N	V	W	Y	H	Y
Cluster	Energy														
1	709.4	Y	Y	P	V	Y	L	Y	T	S	V	W	C	A	A
2	598.0	Y	Y	S	L	Y	L	Y	T	G	V	W	C	A	N
3	598.8	Y	Y	P	V	Y	A	Y	T	G	V	W	C	Q	K
4	585.8	Y	Y	P	L	Y	L	Y	A	N	V	W	C	F	A
5	609.5	Y	Y	P	L	Y	L	Y	T	G	V	W	C	A	A
6	614.5	Y	Y	P	V	Y	L	Y	T	G	V	W	C	V	R
7	598.4	Y	Y	S	V	Y	L	Y	T	G	V	W	C	Q	K
8	608.0	Y	Y	P	L	Y	L	Y	A	G	V	W	C	T	A
9	584.9	Y	Y	P	L	Y	L	Y	A	N	V	W	C	Q	A
10	708.9	Y	Y	S	V	Y	L	Y	T	S	V	W	C	Q	A

Figure 17b

	Chain	L	L	L	L	L	L	L	L	H	H	H	H	H	H
	1CZ8 Position	32	36	43	46	49	50	87	91	35	37	47	95	107	109
	Chothia Position	32	36	43	46	49	50	87	91	35	37	47	91	100c	100e
	Amino Acids Considered	YW DA NS L	YF L	AV SP	VL SR HF	YF	YD AS YK LE TG WQ	Y F	YD LA FH TS GR	NH SQ G	VI AF	WY	YH C	All 20	All 20
	WT	Y	Y	A	V	Y	F	Y	Y	N	V	W	Y	H	Y
Cluster	Energy														
5	609.5	Y	Y	P	L	Y	L	Y	T	G	V	W	C	A	A
5	616.0	Y	Y	S	L	Y	L	Y	A	G	V	W	C	A	S
5	706.4	Y	Y	S	L	Y	L	Y	T	S	V	W	C	A	A
5	706.5	Y	Y	P	L	Y	L	Y	T	S	A	W	C	A	A
5	706.7	N	Y	P	L	Y	L	Y	A	S	A	W	C	A	A
5	707.5	Y	Y	P	L	Y	L	Y	A	S	V	W	C	A	A
5	708.1	Y	Y	S	L	Y	L	Y	A	S	V	W	C	A	S
5	708.4	A	Y	P	L	Y	L	Y	A	S	A	W	C	A	A
5	709.0	Y	Y	P	L	Y	L	Y	A	S	V	W	C	A	A
5	709.4	Y	Y	P	L	Y	L	Y	T	S	V	W	Y	A	A

Figure 18



Figure 19a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
36 L	Core	Y	Y	Y:501 F:449 M:32 L:18
43 L	Boundary	A	Y	Y:481 F:204 H:141 D:42 T:42 A:41 N:28 S:21
87 L	Core	Y	F	F:557 Y:411 M:32
89 L	Core	Q	M	M:1000
98 L	Core	F	F	F:960 Y:40
37 H	Core	V	V	V:1000
45 H	Core	L	L	L:954 M:46
95 H	Core	Y	F	F:613 Y:330 M:49 L:8
107 H	Core	M	M	M:1000
110 H	Core	W	W	W:1000

Figure 19b

Position	Experimental Library
36 L	Y F M L
43 L	Y F H D T A N S
87 L	F Y M
89 L	M Q
98 L	F Y
37 H	V
45 H	L M
95 H	F Y M
107 H	M
110 H	W
Complexity	2304

Figure 19c

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
43 L	Surface	A	H	H:532 T:161 D:93 S:84 N:77 A:53
87 L	Surface	Y	K	K:348 H:248 E:132 R:75 S:68 Q:57 D:40 A:18 T:11 N:3
98 L	Surface	F	H	H:1000
37 H	Surface	V	E	E:546 K:229 T:119 D:35 H:35 Q:30 A:6
45 H	Surface	L	E	E:379 Q:275 T:145 D:48 A:47 R:43 K:31 N:17 S:15
107 H	Surface	M	H	H:460 E:419 Q:47 K:36 D:32 A:6

Figure 19d

Position	Experimental Library
43 L	H T D S N A
87 L	K H E R S Q Y
98 L	H F
37 H	E K T V
45 H	E Q T L
107 H	H E M
Complexity	4032

Figure 20



Figure 21a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
116 L	Core	F	F	F:877 Y:100 M:23
118 L	Core	F	F	F:998 Y:2
133 L	Core	V	M	I:444 M:292 V:264
135 L	Core	L	L	L:790 I:148 M:54 V:8
176 L	Core	S	M	M:428 V:365 A:118 S:89
178 L	Core	T	M	M:910 V:47 I:42 L:1
134 H	Core	L	M	L:709 M:291
147 H	Core	A	A	A:1000
151 H	Core	L	L	L:979 M:21
176 H	Core	F	F	F:982 Y:18
189 H	Core	S	V	V:666 M:277 A:57
191 H	Core	V	V	V:677 M:258 I:65

Figure 21b

Position	Experimental Library
116 L	F Y
118 L	F
133 L	I M V
135 L	L I M
176 L	M V A S
178 L	M T V
134 H	L M
147 H	A
151 H	L
176 H	F
189 H	V M A S
191 H	V M I
Complexity	5184

Figure 22



Figure 23a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
351 A	Core	L	L	L:687 M:304 I:8 A:1
366 A	Core	T	V	V:433 T:251 A:122 M:107 I:87
368 A	Core	L	L	L:899 I: 64 V:30 M:4 A:3
394 A	Core	T	T	T:980 A:20
405 A	Core	F	F	F:966 M:23 Y:11
407 A	Core	Y	F	F:619 M:245 Y:136
351 B	Core	L	L	L:680 M:313 I:6 A:1
366 B	Core	T	V	V:433 T:221 I:129 A:107 M:105 L:5
368 B	Core	L	L	L:927 I:48 V:22 A:3
394 B	Core	T	T	T:978 A:22
405 B	Core	F	F	F:981 Y:10 M:9
407 B	Core	Y	F	F:641 M:204 Y:155

Figure 23b

Position	Experimental Library
351 A	L M
366 A	V T A M I
368 A	L I
394 A	T
405 A	F
407 A	F M Y
351 B	L M
366 B	V T I A M
368 B	L
394 B	T
405 B	F
407 B	F M Y
Complexity	1800

Figure 24



Figure 25a

Position	WT	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
54 L	L								I	K	L			P	Q		S	T			
92 L	I									K											
110 L	V	A		D					I				N	P			S		V		
154 L	L				E						L		N	P	Q	R	S	T	V		
191 L	V										L		N								
73 H	L					F										R					
116 H	L										L			P							
178 H	L			D									N			R					
203 H	I								I	K									V		

Figure 25b

Position	Experimental Library
54 L	L K Q S T
92 L	I K
110 L	V A D N S
154 L	L E N Q R S T
191 L	V N
73 H	L R
116 H	L
178 H	L D N R
203 H	I K
Complexity	11200

Figure 26

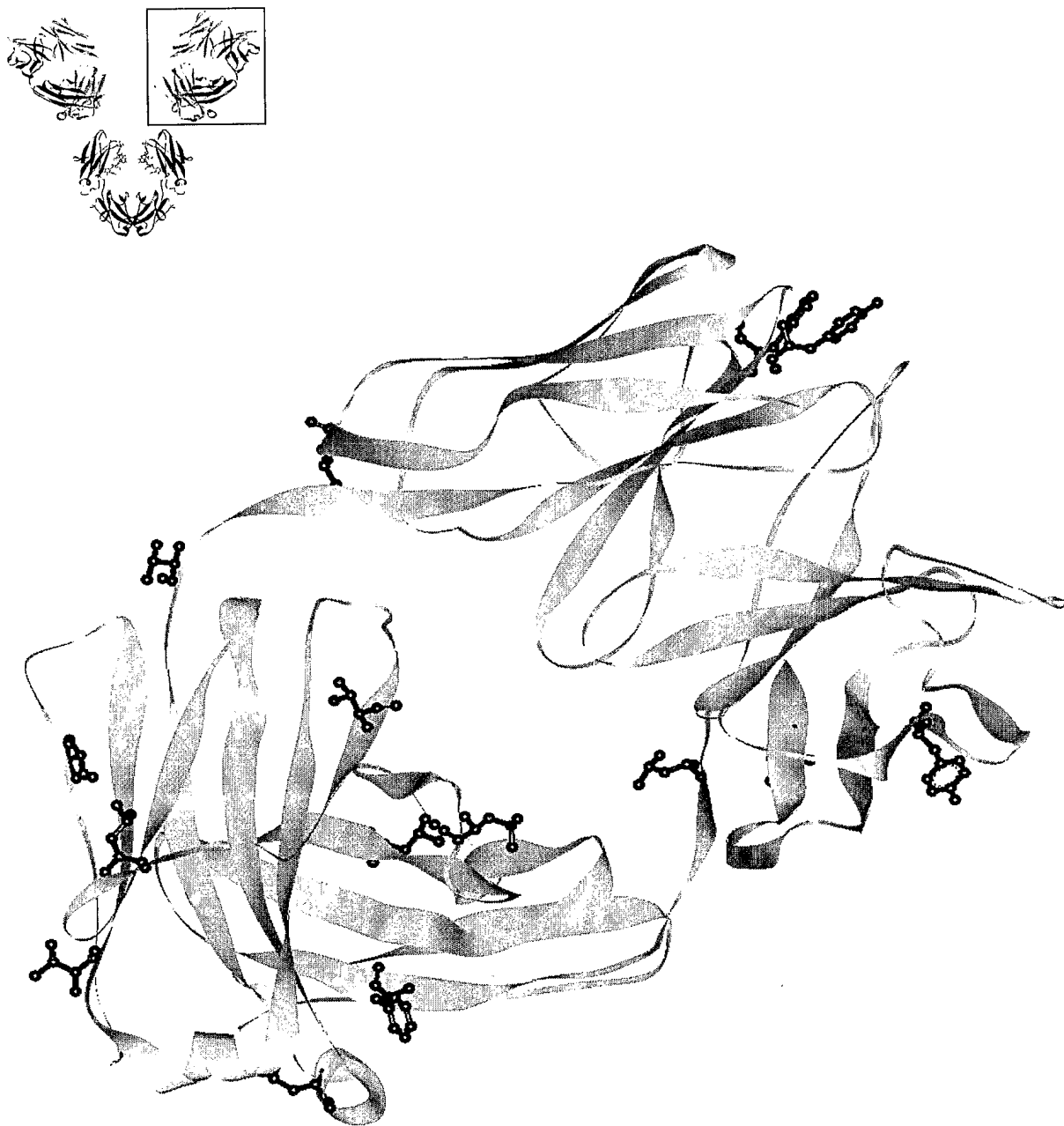


Figure 27a

Position	WT	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
15 L	V								I										V		
32 L	Y															R					
50 L	F	A																			
110 L	V	A		D					I				N	P					V		
125 L	L										L										
154 L	L				E									P	Q		S	T			
163 L	V	A							I								S	T	V		
191 L	V	A								K	L						S		V		
205 L	V														Q						
5 H	V															R					
80 H	Y					F															Y
118 H	L										L										
132 H	F					F										R					
169 H	L										L										
180 H	L			D	E						L		N		Q	R					

Figure 27b

Position	Experimental Library
15 L	V
32 L	Y R
50 L	F A
110 L	V A D N
125 L	L
154 L	L E Q S T
163 L	V A S T
191 L	V A K S
205 L	V Q
5 H	V R
80 H	Y
118 H	L
132 H	F R
169 H	L
180 H	L D E N Q R
Complexity	61440

Figure 28



Figure 29a

Position	WT	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
15 L	V				E				I						Q			T	V		
83 L	F					F															
106 L	I				E				I	K				P	Q	R		T			
11 H	L	A								K			N		Q	R	S				Y
93 H	V	A				F											S				
115 H	L								I				N		Q		S	T	V		

Figure 29b

Position	Experimental Library
15 L	V E Q T
83 L	F
106 L	I E K Q R T
11 H	L A K N Q R S
93 H	V A S
115 H	L N Q S T
Complexity	2520

Figure 30

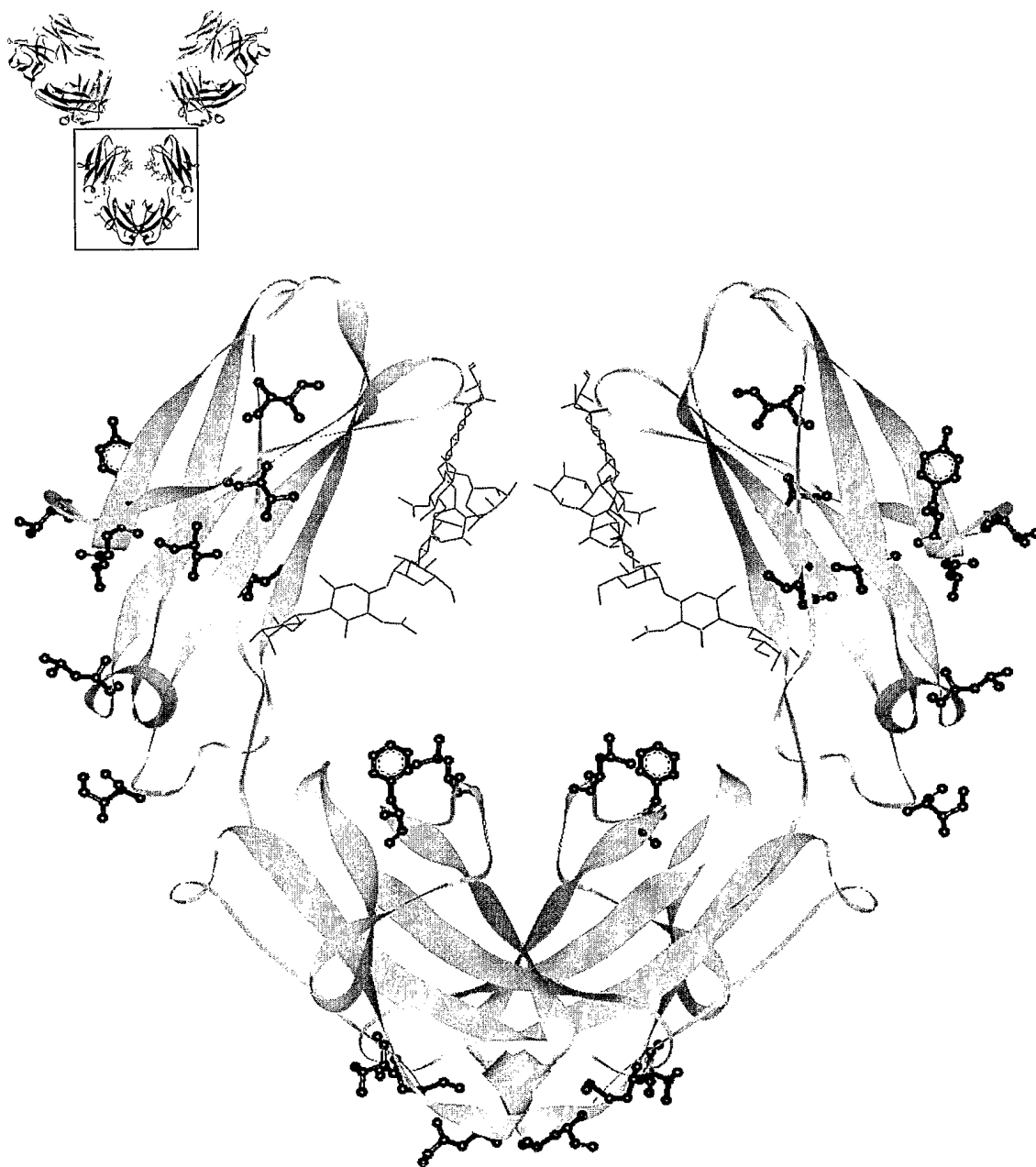


Figure 31a

Position	WT	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
253	I	A								K	L										
278	Y	A			E					K					Q		S	T			Y
279	V								I	K					Q						
282	V	A			E				I	K			N		Q	R	S	T	V		
303	V	A	C		E									P			S	T	V		
305	V								I									T	V		
309	L								I				N	P		R			V		
332	I										L				Q						
336	I	A							I						Q			T	V		
358	M				E					K	L	M			Q		S				
398	L	A		D	E					K	L		N	P	Q	R	S				
404	F					F															Y
422	V	A							I	K	L		N	P	Q	R	S	T	V		
443	L	A		D	E				I	K	L		N		Q	R	S	T			

Figure 31b

Position	Experimental Library																			
253	I	A	K																	
278	Y	A	E	K	Q	S	T													
279	V	K	Q																	
282	V	A	E	K	N	Q	R	S	T											
303	V	A	E	S	T															
305	V	T																		
309	L	N	R																	
332	I	Q																		
336	I	A	Q	T																
358	M	E	K	Q	S															
398	L	A	D	E	K	N	Q	R	S											
404	F																			
422	V	A	K	N	Q	R	S	T												
443	L	A	D	E	K	N	Q	R	S	T										
Complexity	489888000																			

Figure 32

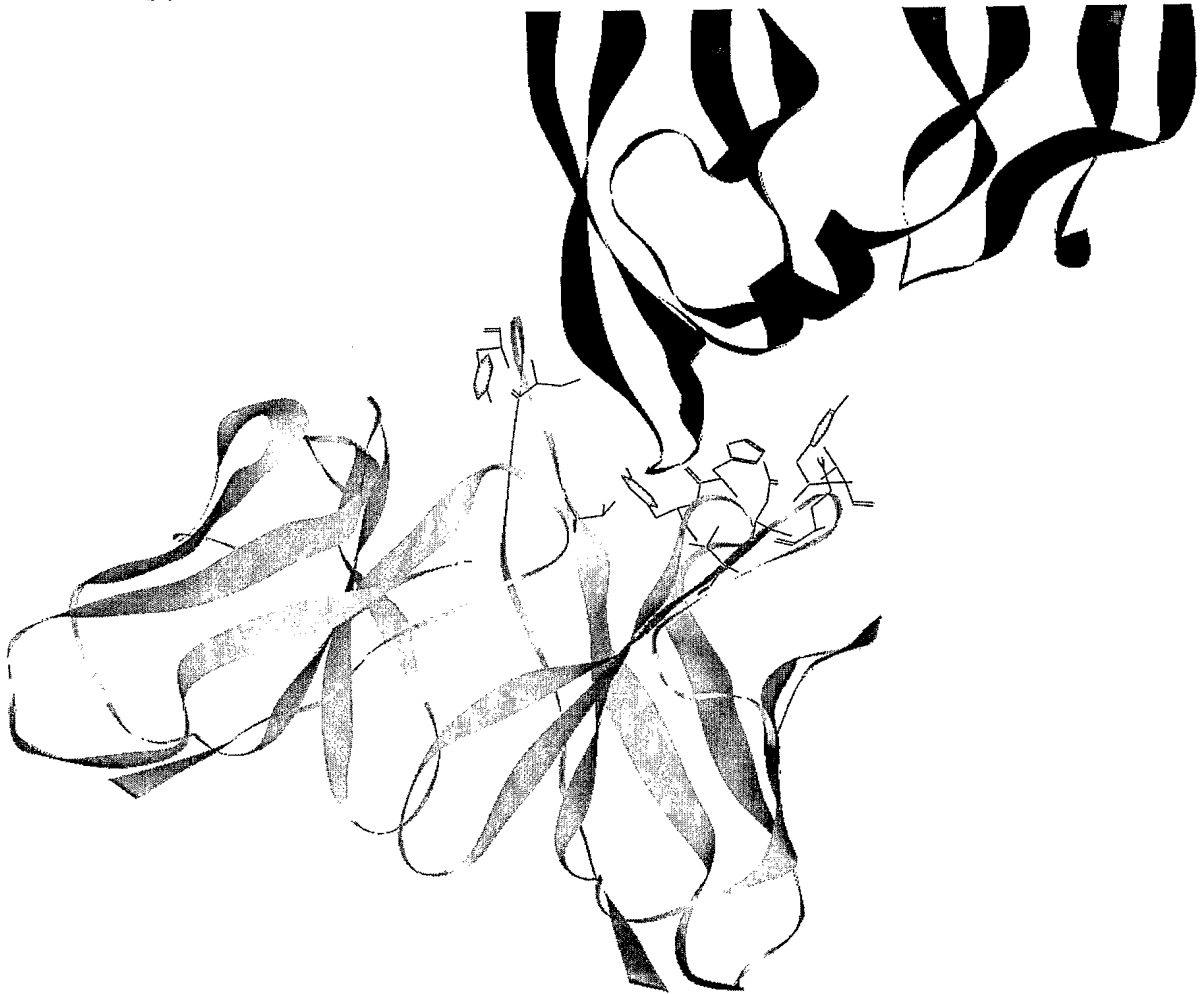
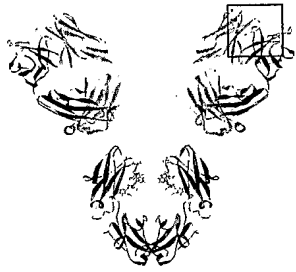


Figure 33a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
94 L	Surface	V	V	V:832 T:149 E:19
31 H	Core	H	W	W:1000
32 H	Core	Y	F	Y:575 F:425
52 H	Boundary	N	N	N:944 D:54 M:2
54 H	Boundary	Y	W	W:878 F:107 Y:15
55 H	Core	T	T	T:994 L:4 M:2
57 H	Core	E	W	W:582 Y:410 E:8
59 H	Surface	T	H	H:472 K:314 E:197 Q:17
103 H	Surface	Y	K	K:508 H:220 Y:201 R:68 T:3
106 H	Surface	S	T	T:692 S:308

Figure 33b

Position	Experimental Library
94 L	V T
31 H	W H
32 H	Y F
52 H	N D
54 H	W F Y
55 H	T
57 H	W Y E
59 H	H K E T
103 H	K H Y R
106 H	T S
Complexity	2304

Figure 34



Figure 35a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
91 L	Boundary	Y	F	F:579 Y:407 W:9 M:4 H:1
27 H	Core	Y	F	F:521 Y:475 M:4
28 H	Surface	D	D	D:327 T:185 E:154 Q:86 K:65 S:55 N:51 R:47 H:20 A:10
30 H	Surface	T	R	T:230 R:210 E:137 K:121 Q:108 H:96 S:56 A:19 D:14 N:9
35 H	Core	N	N	N:826 A:173 V:1
47 H	Core	W	W	W:1000
98 H	Boundary	K	K	K:913 M:87
110 H	Core	F	F	F:966 L:26 M:8

Figure 35b

Position	Experimental Library
91 L	F Y
27 H	F Y
28 H	D T E Q K S N
30 H	T R E K Q H S
35 H	N A
47 H	W
98 H	K M
110 H	F
Complexity	784

Figure 36



Figure 37a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
32 L	Core	Y	Y	Y:807 F:154 M:39
50 L	Boundary	G	K	K:410 M:282 E:77 V:67 T:62 Y:34 R:21 Q:20 D:10 F:5 S:5 A:4 N:3
53 L	Boundary	N	Q	Q:857 K:66 Y:25 I:13 W:11 E:7 M:6 R:4 H:4 V:3 F:2 L:2
56 L	Boundary	P	T	T:446 R:156 S:87 K:82 Q:69 E:60 V:50 M:25 N:9 H:9 A:5 Y:2
93 L	Boundary	S	E	E:478 Y:127 T:115 D:84 Q:81 S:42 N:38 K:22 I:13
96 H	Boundary	G	L	L:573 M:417 V:6 K:4
97 H	Surface	Q	Q	E:517 Q:483

Figure 37b

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
27 H	Core	F	F	F:834 Y:136 M:20 L:10
31 H	Boundary	N	W	W:208 N:152 D:148 E:146 A:106 Q:88 S:59 R:38 M:32 K:23
32 H	Boundary	Y	Y	Y:698 F:302
33 H	Boundary	W	W	W:1000
52 H	Boundary	R	L	L:447 R:389 M:66 K:66 W:13 T:12 E:6 Q:1
95 H	Boundary	V	V	V:947 T:43 M:6 A:4
102 H	Core	Y	Y	Y:699 W:157 F:143 A:1

Figure 37c

Position	WT	Experimental Library
32 L	Y	Y F
50 L	G	K M E V T G
53 L	N	Q K N
56 L	P	T R S K Q E V P
93 L	S	E Y T D Q S
27 H	F	F Y
31 H	N	W N D E A Q S
32 H	Y	Y F
33 H	W	W
52 H	R	L R M K
95 H	V	V
96 H	G	L M G
97 H	Q	E Q
102 H	Y	Y W F
Complexity		3483648

Figure 38



Figure 39a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
32 L	Core	Y	Y	Y:1000
50 L	Boundary	N	M	M:982 Q:18
91 L	Surface	H	H	H:1000
94 L	Surface	R	K	R:454 K:299 T:238 E:5 Q:2 S 2
50 H	Core	F	F	F:987 Y:13
52 H	Surface	R	R	R:1000
103 H	Surface	H	H	H:516 D:128 T:115 K:96 S:71 E:22 R:20 N:16 Q:16
106 H	Boundary	A	S	S:656 A:263 T:81

Figure 39b

Position	Sequences Around Ground State
32 L	Y
50 L	M Q N
91 L	H
94 L	R K T
50 H	F Y
52 H	R
103 H	H D T K S E R N Q
106 H	S A T
Complexity	486

**Figure 40a**

[illegible]

[illegible]

[illegible]

[illegible]

Figure 41a

Chain		L	L	L	L	H	H	H	H	H	H
1CE1 Position		32	50	91	94	33	50	52	56	103	106
Chothia Position		32	50	91	94	33	50	52	53	97	100
Amino Acids Considered		YW DA NS L	ND AS YK LE TG KW Q	HY DL AF TS GR	RY FA LT WS H	YF AD GS CW ET	FR WG IL HN YA SV E	RN SD TV IY FK GW	KN GY EF SG IW DA H	All	All
WT		Y	N	H	R	Y	F	R	K	H	A
Cluster	Energy										
1	477.7	A	K	T	L	A	L	R	K	S	S
2	472.8	A	K	T	L	E	F	S	K	S	S
3	477.5	A	K	T	T	A	F	R	K	S	S
4	471.6	D	K	H	L	A	L	R	K	S	A
5	469.6	Y	K	F	L	E	F	K	K	S	S
6	471.7	D	K	F	L	E	F	S	K	S	A
7	467.1	Y	K	H	L	A	F	R	K	S	S
8	466.7	Y	K	H	L	S	L	K	N	S	S
9	472.5	A	K	T	L	E	F	K	K	S	S
10	465.5	Y	K	H	L	E	F	K	K	S	S

Figure 41b

Chain		L	L	L	L	H	H	H	H	H	H
1CE1 Position		32	50	91	94	33	50	52	56	103	106
Chothia Position		32	50	91	94	33	50	52	53	97	100
Amino Acids Considered		YW DA NS L	ND AS YK LE TG KW Q	HY DL AF TS GR	RY FA LT WS H	YF AD GS CW ET	FR WG IL HN YA SV E	RN SD TV IY FK GW	KN GY EF SG IW DA H	All 20	All 20
WT		Y	N	H	R	Y	F	R	K	H	A
Cluster	Energy										
4	471.6	D	K	H	L	A	L	R	K	S	A
4	471.7	A	L	H	L	A	L	R	K	S	S
4	472.6	D	K	H	L	A	L	R	K	S	A
4	473.4	Y	K	F	L	A	L	R	K	S	A
4	475.9	Y	K	F	T	A	L	R	K	S	A
4	476.1	N	K	H	R	A	L	K	N	S	A
9	472.5	A	K	T	L	E	F	K	K	S	S
9	476.2	A	K	T	L	S	L	K	S	S	A
9	477.2	A	K	T	R	S	L	K	N	S	A
9	478.6	A	K	T	L	S	F	K	K	S	A
9	481.5	A	K	T	L	S	L	K	S	A	S

Figure 42



Figure 43a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
94 L	Boundary	S	S	S:974 A:26
32 H	Boundary	Y	F	F:780 Y:220
50 H	Core	L	L	L:1000
55 H	Surface	Q	Q	Q:776 R:207 K:17
57 H	Core	N	W	W:709 M:109 F:83 Y:68 I:30 N:1
59 H	Boundary	I	I	I:1000
65 H	Surface	Q	K	K:293 R:258 H:224 Q:72 E:42 T:34 N:32 D:23 S:17 A:5
101 H	Boundary	A	A	A:942 S:52 D:6

Figure 43b

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
94 L	Boundary	S	F	F:793 Y:113 M:60 W:13 H:8 I:4 E:3 K:3 Q:2 V:1
32 H	Boundary	Y	Y	Y:907 F:92 H:1
50 H	Core	L	M	M:802 L:198
55 H	Surface	Q	H	H:429 K:196 Q:152 R:139 T:29 N:21 E:18 S:8 D:5 A:3
57 H	Core	N	Y	N:523 Y:432 A:25 V:7 L:5 F:2 I:2 M:2 W:2
59 H	Boundary	I	E	E:565 I:151 Y:127 Q:102 K:17 V:16 R:6 T:6 W:3 F:2 L:2 M:2 H:1
65 H	Surface	Q	R	R:880 K:41 H:20 Q:18 T:18 N:11 E:7 S:5
101 H	Boundary	A	K	K:504 Q:85 R:81 N:75 H:72 S:49 Y:28 E:22 T:16 D:15 I:15 F:12 M:10 A:7 L:7 V:1 W:1

Figure 43c

Position	Bound Experimental Library	Unbound Experimental Library
94 L	S A	F Y M W S
32 H	F Y	Y F
50 H	L	M L
55 H	Q R K	H K Q R T N E
57 H	W M F Y I N	N Y A
59 H	I	E I Y Q K V
65 H	K R H Q E T N D S	R K H Q T N
101 H	A S	K Q R N H S Y E T D I F M A
Complexity	1296	211680

Figure 43d

Position	Experimental Library
94 L	S
32 H	F Y
50 H	L
55 H	Q R K
57 H	W M F Y I N
59 H	I
65 H	K R H Q E T N S
101 H	A S D
Complexity	864

Figure 44



Figure 45a

Position	Considered Amino Acids	WT	Ground State	Sequences Around Ground State
30 L	Surface	N	S	S:519 D:125 N:104 E:78 K:66 Q:64 H:28 R:10 T:6
49 L	Core	Y	Y	Y:1000
53 L	Core	F	Y	Y:1000
55 L	Core	Y	Y	Y:579 F:421
91 L	Boundary	H	W	W:687 F:313
92 L	Core	Y	Y	Y:1000
94 L	Boundary	T	E	V:511 E:196 Y:150 I:129 T:10 Q:2 W:2
33 H	Core	Y	Y	Y:999 F:1
57 H	Core	Y	Y	Y:397 I:331 W:143 L:63 V:60 A:6
59 H	Boundary	R	K	Y:446 Q:311 K:173 E:26 M:23 L:21
104 H	Core	F	F	F:1000

Figure 45b

Position	Experimental Library
30 L	S D N E K Q H R
49 L	Y
53 L	Y F
55 L	Y F
91 L	W F H
92 L	Y
94 L	V E Y I T
33 H	Y
57 H	Y I W L V
59 H	Y Q K E M L R
104 H	F
Complexity	16800